# Conclusive Exclusion of Quantum States and Aspects of Thermo-Majorization

**By**

**Christopher David Perry**

*A thesis submitted to*

University College London

*for the degree of*

Doctor of Philosophy

Department of Physics and Astronomy

University College London

$30^{\text{th}}$ December, 2015

I, Christopher David Perry, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed ..... Date .....

**Conclusive Exclusion of Quantum States and Aspects of Thermo-Majorization**

**By**

**Christopher David Perry**

Doctor of Philosophy of Physics

University College London

Prof. Jonathan Oppenheim, Supervisor

## Abstract

**Part 1**: Why can we not distinguish between pure, non-orthogonal quantum states? Regarding the quantum state as a state of knowledge rather than something physically real, has the potential to answer this question and explain other quantum properties but such interpretations have recently been undermined. This important no-go result, due to Pusey, Barrett and Rudolph, makes use of a specific example of a task we term *state exclusion*. Here, a system is prepared in a state chosen from a known set and the aim is to determine a preparation that has not taken place. We formulate state exclusion as a semidefinite program, using it to investigate when exclusion is conclusively possible and how it can be achieved.

Based on state exclusion, we construct a communication task which exhibits drastic, '*infinite*', separations between a variety of classical and quantum information and communication complexity measures. This serves to requisition the aforementioned foundational result for use in information theoretic protocols.

**Part 2**: What does thermodynamics look like in the absence of the thermodynamic limit? In recent years there has been a concerted effort to apply techniques from quantum information theory to study the laws of thermodynamics at the nano-scale. This has led to the resource theory of *thermal operations* for determining when single-copy transformations are possible. However, if a deterministic transition is forbidden, can it occur probabilistically? Here we compute and bound the maximum probability with which nano-scale thermodynamical transformations can occur.

Thermal operations assume that one can precisely manipulate all of the degrees of freedom in a very large heat bath. While this enables the derivation of ultimate limits on nano-scale

thermodynamics, it does not make them feasible to perform in reality. We show how allowed transitions can be implemented whilst manipulating only a single bath-qubit, making thermal operations more experimentally palatable.

# List of Publications and Preprints

The majority of the work presented in this thesis contains materials from the following publications:

S. Bandyopadhyay, R. Jain, J. Oppenheim, and C. Perry, *Conclusive exclusion of quantum states*, Phys. Rev. A, **89**, 022336 (2014).

T.A. Brun, M.-H. Hsieh, and C. Perry, *Compatibility of state assignments and pooling of information*, Phys. Rev. A, **92**, 012107 (2015).

C. Perry, R. Jain, and J. Oppenheim, *Communication tasks with infinite quantum-classical separation*, Phys. Rev. Lett. **115**, 030504 (2015).

Á.M. Alhambra, J. Oppenheim, and C. Perry, *What is the probability of a thermodynamical transition?*, Preprint arXiv:1504.00020.

Z.-W. Liu, C. Perry, Y. Zhu, D.E. Koh, and S. Aaronson, *Doubly infinite separation of quantum information and communication*, Preprint arXiv:1507.03546.

C. Perry, P. Ćwikliński, J. Anders, M. Horodecki, and J. Oppenheim, *A sufficient set of experimentally implementable thermal operations*, Preprint arXiv:1511.06553.

# Acknowledgments

The three years of my PhD and the completion of this thesis would have been near impossible without the input of many people both from a physics and sanity-preservation perspective. On the physics front, the foremost contributor has been my supervisor Jonathan. I am grateful to him for plucking me from a career of Excel spreadsheets in Stevenage to undertake the research presented here, for teaching me a lot of physics and for finally convincing me to take the plunge and learn a little thermodynamics. I am less grateful for the amount of chalk dust sitting in my lungs from his office's blackboard-wall combination.

I have been fortunate to have had the opportunity to accumulate a number of excellent collaborators. In particular, I thank Som for his encouragement - especially in the early years, Rahul for his patience in guiding me through the subtleties of communication complexity theory and Álvaro for his willingness to decorate every whiteboard in the department with scrawled Lorenz curves. I have also benefited from many conversations with people both at UCL and on my travels and I cannot do them all justice here. Special thanks however, goes to Neil, and Naïri and Hussain for countless 'discussions' (particularly to Hussain for undertaking the task of proofreading this thesis for me). My examiners, Toby Cubitt and Tony Short, also deserve thanks: firstly for taking the time to read this tomb and secondly for their insightful comments that have helped to improve it.

Sanity-preservation duties have been ably shouldered by my original office-mates, Pete and Naïri. $K_3$ drinks, flippantry, Jamie Dalrymple's profligacy, extended afternoon strolls and much more besides have made PhD life a lot easier. It would also be remiss of me not to mention the 12:15 lunch crew. Arne, Alexandros, The Moderate Johnny, Rafa, Roberta, Sarah, Álvaro, Enrico, Luca, Pik, Brendan and the adopted Brummies, İlhan and Michela, have provided a welcome daily escape from my desk.

Thank you to my parents and brother for all of their love, support and encouragement

throughout my life. Without it, I definitely would not have made it to where I am today. Finally, my deepest thanks go to Marie. Her smiles, laughter and love are invaluable.

# Contents

# List of Figures

# Introduction

*The devil speed him! No man's pie is freed*
*From his ambitious finger.*

- The Duke of Buckingham, regarding Cardinal Wolsey in Shakespeare's *Henry VIII*

Much like Cardinal Wolsey, quantum information theory has its fingers in many pies. With a broad approach of applying concepts from information science to investigate the laws of physics, few topics lie beyond its reach. It has been used for tasks as diverse as exploring the underlying foundations of quantum theory, guiding the development of fundamental laws for small-scale thermodynamics and studying the power of computation both within the quantum realm and beyond.

Often the insights that quantum information theory brings is to phrase problems and concepts as a question of resources. Given resources from a particular physical theory, what can we achieve? Alternatively, for a given task, how well can we perform it if we are restricted to using resources governed by classical physics, quantum physics or some more general (or more restrictive) set? In this resource theoretic approach, we assign a notion of value to systems based on our ability to manipulate them: a system we have to work hard to produce yet vital for reaching some goal is intrinsically more valuable than another we can create with little effort. Similarly, by identifying tasks that we believe should be difficult, we can cast aside models of physics in which they are too easy or, in the face of insurmountable evidence supporting such theories, adjust our intuition accordingly and attempt to exploit them.

This thesis investigates two areas in which this style of questioning bears fruit. Firstly we consider a central task within quantum information theory: investigating the advantages in using quantum resources over classical ones. For example, in communication tasks, two players are interested in computing a function of their inputs and must exchange messages to do so.

What are the benefits in using quantum over classical messages? The main result of Part I is the development of a communication task, *the exclusion game* [140, 113], for which sending quantum messages is dramatically more powerful than sending classical ones. Considering the amount of information the messages must contain regarding their inputs leads to an 'infinite' separation between the power of quantum and classical resources. Classically the players must reveal nearly everything about their inputs while a quantum strategy can succeed and yet reveal next to nothing.

Furthermore, a variation on this task in which a quantum strategy has access to shared entanglement also leads to striking results. Such a strategy exists which needs only a constant number of bits to be exchanged while any purely classical procedure requires that practically the entirety of the input be sent. Before this work, only exponential separations were known for both this and the previous scenarios. In addition, when comparing the amount of communication and information that must be exchanged in a quantum strategy we find separations qualitatively different to those known in the classical setting.

In analyzing this communication task, we define and develop the problem of *state exclusion* [10]. There, we consider a quantum system prepared in a state chosen from a known set. The aim is to perform a measurement on the system which can conclusively rule that a subset of preparations cannot have taken place. By formulating state exclusion as a semidefinite program (SDP), we derive necessary and sufficient conditions for an exclusion measurement to be optimal and a necessary condition on the set of states for exclusion to be achievable with certainty. This task of excluding states has recently gained importance in the context of the foundations of quantum mechanics due to a result by Pusey, Barrett and Rudolph (PBR) [142]. Using our SDP we prove the optimality of the PBR proof and this provides part of the quantum strategy for the exclusion game.

Part II of this thesis considers the laws of thermodynamics for nano-scale systems. While traditional thermodynamics considers the macroscopic properties of systems composed of many particles, nano-scale thermodynamics studies the microscopic degrees of freedom of single systems. In essence, this is thermodynamics in the absence of the thermodynamic limit. Within this regime, recent work applying techniques developed in the context of quantum information theory has led to the derivation of necessary and sufficient conditions for thermodynamical transformations to be possible - a plethora of second laws [89, 24].

If these laws forbid a transition between two states, it can still be made possible provided a sufficient amount of work is supplied. Suppose however, that we cannot or do not wish to expend this work. Can the transition occur probabilistically rather than with certainty? In the thermodynamic limit, the answer is no as the probability tends to zero. However, here we find that for finite-sized systems it can be non-zero [4], calculating the maximum probability for a transition between any initial and final states and showing that this maximum can be achieved when the target state is block-diagonal in the energy eigenbasis. Furthermore, the probability of a transition is intrinsically related to the amount of work that would be required to drive the transformation with certainty and we develop new methods for calculating this quantity.

The final contribution of this thesis is to make experimental test of the laws of nano-scale thermodynamics more feasible [139]. In deriving the aforementioned second laws, it is assumed that one has precise control over the microscopic degrees of freedom of not only the system under consideration but also an extremely large heat bath. Allowing for such fine-grained manipulation still leads to constraints on which state transformations are possible and enables one to implement these transitions. However, performing such a specific operation on joint micro-states of the system and bath is beyond the reach of current experiments. Surprisingly, we show here that the ability to apply two simple operations to a system combined with a single thermal qubit allows one to carry out any transformation that can be performed under the second laws. This serves to make thermal operations more experimentally accessible and raises the possibility of performing thermodynamical processes that have never before been seen in the laboratory.

Throughout this thesis, we shall tend to assume that the reader is familiar with the basics of quantum mechanics and quantum information theory. Excellent introductions to these topics can be found, for example, in [138, 134, 169].

# Part I

# Conclusive Exclusion of Quantum States

# Guide to Part I

We begin with Chapter 1, which revolves around one of the most basic and well studied tasks in quantum information theory: that of state discrimination. Here we review how state discrimination can be cast as an optimization problem, a semidefinite program (SDP), and highlight the insights such a formulation provides.

Our reasons for opening with a discussion of state discrimination are twofold. Firstly, instead of analyzing how well one can distinguish between states, one could instead ask the question as to why perfect pure state discrimination is impossible in quantum theory? The framework of ontological models provides a structure to attempt to answer this question as well as to explain other curious features of quantum theory. Chapter 2 begins with a discussion of such constructions, before leading into the recent theorem of Pusey, Barrett and Rudolph (PBR) [142] regarding the reality of the quantum state which puts severe constraints on the form these models can take. The proof of the PBR theorem revolves around an instance of a task we christen state exclusion. In this task, given a system prepared from a set of possible states, rather than attempting to identify which state has been prepared, one tries to determine a state that has not.

This brings us to our second reason behind the subject matter of Chapter 1. Closely related to state discrimination, state exclusion can also be formulated as an SDP. The second half of Chapter 2 explores this optimization problem, using it to determine:

- Necessary and sufficient conditions for a given measurement to be optimal for attempting state exclusion.
- Conditions for state exclusion to be possible.
- Lower bounds on the probability of making an error.

On top of this, the SDP enables us to prove the optimality of a measurement used in the PBR

proof. These are the results contained in [10], joint work with Somshubhro Bandyopadhyay, Rahul Jain and Jonathan Oppenheim. Our SDP also forms part of a set that can be used to quantify how compatible states assignments are according to a hierarchy of criteria. These constructions appeared in work with Todd Brun and Min-Hsiu Hsieh [29].

Part I closes with Chapter 3. Here, based on the PBR theorem and state exclusion, we define a communication task: the exclusion game. In communication tasks, two parties are typically interested in computing a function of their inputs and there exist problems for which this can be done using exponentially less communication by using quantum rather than classical messages. Indeed, for this measure (the communication complexity), it is known that an exponential separation is the best that can be achieved for bounded error.

The exclusion game however, exhibits a significantly larger separation between the quantum and the classical in two scenarios. Firstly, rather than analyzing how much communication is required in a given task, one can consider the amount of information regarding the players' inputs that needs to be revealed. For the exclusion game, it is possible to have an infinite separation with respect to this measure (the information complexity): classically nearly all the information needs to be revealed, while a quantum strategy can succeed and reveal next to nothing. Previously, only an exponential separation was known.

Secondly, if we allow a quantum strategy to have access to entanglement and the players to abort the game with constant probability, it is possible to gain a similar separations with respect to the communication complexity. A quantum strategy exists which requires only a constant number of bits to be exchanged, while any classical strategy requires practically everything to be sent. Again, before this work, only exponential separations had been found.

Furthermore, the game exhibits an infinite separation between the quantum information and communication complexities. Such a separation is qualitatively different to those known previously for the analogous classical complexities.

These exclusion game results can be found in two papers, the first with Rahul Jain and Jonathan Oppenheim [140] and the second with Zi-Wen Liu, Yechao Zhu, Dax Enshan Koh and Scott Aaronson [113].

# Chapter 1

# State Discrimination

## 1.1 The task of state discrimination

One of the defining features that separates quantum mechanics from its classical counterpart is the fact that within its framework, there exists pure states that cannot be distinguished from each other perfectly. That is, if we are given a system prepared in some unknown state, $|\phi\rangle$, chosen at random from a set of two distinct known pure states, $|\psi_1\rangle$ or $|\psi_2\rangle$, there may not exist a measurement that we can perform on the system which allows us to identify which state we were given. In fact, it is a well known result that quantum states (pure or mixed) can be perfectly distinguished from one another if and only if they are orthogonal [134].

This leads to the following natural question: how well can a given set of states be distinguished from one another? That is, given an unknown state, $\sigma$, selected according to some probability distribution, $\{p_i\}_{i=1}^k$, from a set of $k$ known states, $\mathcal{P} = \{\rho_i\}_{i=1}^k$, how accurately can we determine which of the $\rho_i$ was prepared and what is the *optimal* measurement for doing this?

Answering this question is of fundamental importance in much of quantum information theory. At its most basic level we can imagine a communication task in which Alice wishes to transmit a single bit to Bob but she is restricted (perhaps by some noise in the communication channel the two parties share) to encoding this in one of two quantum states, $\rho_0$ and $\rho_1$. The distinguishability of $\rho_0$ and $\rho_1$ determines Bob's ability to successfully determine Alice's message. Conversely, the inability to perfectly distinguish between non-orthogonal states can provide an advantage. In cryptographic protocols such as B92 [21], knowing how successful an eavesdropper can be in distinguishing between particular states can be used to determine the

level of noise parties should tolerate when implementing the protocol.

There exist many different measures for characterizing the success of a discrimination measurement. In *minimum error state discrimination*, we attempt to minimize the average probability of making an error in identifying the state [84]. Alternatively, we could construct a measurement which allows an inconclusive result, the probability of which we try to minimize, while all other results allow us to identify the prepared state with certainty [91, 57, 137]. This formulation is referred to as *unambiguous state discrimination*. Further variants may interpolate between minimum error and unambiguous discrimination by allowing only a fixed probability of obtaining an inconclusive result [42, 175, 67] or attempt to minimize the *worst-case error* [106].

In this thesis, we shall mainly be concerned with tasks closely related to minimum error state discrimination. Here, the measurement consists of $k$ outcomes, one for each element of $\mathcal{P}$, and when outcome $j$ is observed, it is declared that $\sigma = \rho_j$. We are interested in minimizing the probability of error, or equivalently maximizing the probability of success given by:

$$p_{\text{succ}} = \sum_{i=1}^{k} p_i \text{Tr}\left[\rho_i M_i\right], \tag{1.1}$$

over all possible measurements, $\mathcal{M} = \{M_i\}_{i=1}^{k}$. We denote a measurement that achieves the maximum probability of success by $\mathcal{M}^{\text{opt}}$.

This optimization has been widely studied (see [15] for a recent survey). In particular, $\mathcal{M}^{\text{opt}}$ is known for $k = 2$ [84] and for sets of states exhibiting some degree of symmetry [8, 62, 14, 5, 44, 64, 43]. For general sets of states, necessary and sufficient conditions that must be satisfied by $\mathcal{M}^{\text{opt}}$ have been derived [84, 85, 174]. Furthermore, a number of upper and lower bounds on $p_{\text{succ}}$ for given $\{p_i\}_{i=1}^{k}$ and $\mathcal{P}$ have been calculated [16, 124, 125, 143, 144]. Finally, semidefinite programs (SDPs) can be used to give efficient numerical estimates for $p_{\text{succ}}$ [98, 63].

Quantifying how successful it is possible to be in identifying a state from a given set is just one question that arises from the indistinguishability inherent in quantum mechanics. A second, in some ways deeper, question is to ask what gives rise to this property? An attempt to answer this question lies in the framework of ontological models. These will be discussed in Chapter 2 and lead onto the theorem of Pusey, Barrett and Rudolph regarding the reality of the quantum state [142]. This will bring us into the task of state exclusion, closely related to that of state discrimination. State exclusion can also be formulated in the framework of the

aforementioned SDPs and so the rest of this chapter is spent introducing them and exhibiting their utility in the context of state discrimination.

## 1.2 Semidefinite programs

Semidefinite programs are a class of constrained optimization problems that are efficiently numerically solvable and possess a structure that can be exploited to derive statements about the underlying problem they describe [171]. As has already been noted, state discrimination can be formulated as an SDP and within quantum information theory they have been applied to a wide range of problems including (but certainly not limited to) determining Tsirelson-type bounds on Bell inequalities [167], characterizing the set of separable density operators [58, 22] and bounding the capacity of graphs [59]. We now proceed to give a brief introduction to the semidefinite programming formalism as found in [166], together with some of their useful properties.

### 1.2.1 The SDP formalism

Given a complex Euclidean space, $\mathcal{X}$, let $L(\mathcal{X})$ denote the set of matrices associated with linear maps that take $\mathcal{X}$ to itself. We define $Herm(\mathcal{X})$ to be the set of matrices within $L(\mathcal{X})$ that are Hermitian.

A semidefinite program is then defined by three elements, $\{A, B, \Phi\}$. Here, $A$ and $B$ are Hermitian matrices, $A \in Herm(\mathcal{X})$ and $B \in Herm(\mathcal{Y})$, where $\mathcal{X}$ and $\mathcal{Y}$ are complex Euclidean spaces. The third element, $\Phi$, is a Hermicity preserving super-operator which take elements in $Herm(\mathcal{X})$ to elements in $Herm(\mathcal{Y})$.

From these, two optimization problems can be defined: the *primal* and the *dual*. The primal problem is defined to be:

$$\begin{aligned} \underset{X \in L(\mathcal{X})}{\text{Maximize:}} \quad & \alpha = \mathrm{Tr}\left[AX\right]. \\ \text{Subject to:} \quad & \Phi(X) = B, \\ & X \geq 0. \end{aligned} \tag{1.2}$$

The related dual problem is:

$$\begin{aligned} \underset{Y \in L(\mathcal{Y})}{\text{Minimize:}} \quad & \beta = \mathrm{Tr}\left[BY\right]. \\ \text{Subject to:} \quad & \Phi^*\left(Y\right) \geq A, \\ & Y \in \mathit{Herm}\left(\mathcal{Y}\right). \end{aligned} \tag{1.3}$$

Here, $X \geq 0$ implies that $X$ is a positive semidefinite matrix and $\Phi^*$ is the dual map to $\Phi$ defined by:

$$\mathrm{Tr}\left[Y\Phi\left(X\right)\right] = \mathrm{Tr}\left[X\Phi^*\left(Y\right)\right]. \tag{1.4}$$

We shall denote the set of feasible solutions to the primal problem (those $X$ that satisfy the constraints in Eq. (1.2)) by $\mathcal{A}$. Similarly, $\mathcal{B}$ denotes the set of feasible solutions to the dual problem. The optimal value of the primal problem is defined to be:

$$\alpha^{\mathrm{opt}} = \sup_{X \in \mathcal{A}} \mathrm{Tr}\left[AX\right], \tag{1.5}$$

and the optimal value of the dual problem is:

$$\beta^{\mathrm{opt}} = \inf_{Y \in \mathcal{B}} \mathrm{Tr}\left[BY\right]. \tag{1.6}$$

If there exists an $X \in \mathcal{A}$ such that $\alpha^{\mathrm{opt}} = \mathrm{Tr}\left[AX\right]$, then we shall refer to it as an optimal primal solution and denote it by $X^{\mathrm{opt}}$. For the dual problem, $Y^{\mathrm{opt}}$ shall be used to denote a similarly defined optimal dual solution.[1] By convention, if there does not exist an $X$ satisfying the constraints of the primal problem ($\mathcal{A} = \emptyset$), we take $\alpha^{\mathrm{opt}} = -\infty$. If there does not exist a $Y$ satisfying the constraints of the dual problem ($\mathcal{B} = \emptyset$), we take $\beta^{\mathrm{opt}} = \infty$.

### 1.2.2 Properties of SDPs

Part of the power and usefulness of the SDP formalism lies in the relationships between $\alpha^{\mathrm{opt}}$ and $\beta^{\mathrm{opt}}$. These are expressed in the concepts of *weak* and *strong duality*.

[1] Note that $X^{\mathrm{opt}}$ need not be unique. For example, the SDP:

$$\begin{aligned} \underset{X \in L(\mathcal{X})}{\text{Maximize:}} \quad & \mathrm{Tr}\left[X\right]. \\ \text{Subject to:} \quad & \mathrm{Tr}\left[X\right] = 1, \\ & X \geq 0, \end{aligned}$$

obviously has many optimal solutions (all achieving $\alpha^{\mathrm{opt}} = 1$).

### Weak duality

**Proposition 1** (Weak duality for SDPs)**.** *For every semidefinite program,* $\{A, B, \Phi\}$*:*

$$\alpha^{\text{opt}} \leq \beta^{\text{opt}}. \tag{1.7}$$

*Proof.* If $\mathcal{A} = \emptyset$ or $\mathcal{B} = \emptyset$, then $\alpha^{\text{opt}} = -\infty$ or $\beta^{\text{opt}} = \infty$ and the statement obviously holds. Let $X \in \mathcal{A}$ and $Y \in \mathcal{B}$ be feasible solutions to the primal and dual problem. Then:

$$\text{Tr}\left[AX\right] \leq \text{Tr}\left[\Phi^*\left(Y\right)X\right] = \text{Tr}\left[\Phi\left(X\right)Y\right] = \text{Tr}\left[BY\right]. \tag{1.8}$$

Taking the supremum over $X \in \mathcal{A}$ and the infimum over $Y \in \mathcal{B}$ then gives the result. □

Weak duality allows us to place bounds on the optimal value of an SDP. If we can find a $Y \in \mathcal{B}$, then we can infer that $\alpha^{\text{opt}} \leq \text{Tr}\left[BY\right]$ and we have obtained an upper bound on the optimal value of the primal problem. Conversely, $X \in \mathcal{A}$ imply bounds on the optimal value of the dual problem. In Section 1.3.3 we shall see how this can be used to give simple proofs of some known bounds on the probability of successful state discrimination.

### Strong duality

If an SDP is such that equality holds in Eq. (1.7), then it is said to satisfy strong duality. Whilst weak duality holds for every SDP, the same is not true of strong duality though in practice this is often the case. If strong duality does hold, then this gives us a method for checking whether a feasible solution to the primal problem, $X$, is in fact an optimal one. Namely, if we can find a feasible solution to the dual problem, $Y \in \mathcal{B}$, such that $\text{Tr}\left[BY\right] = \text{Tr}\left[AX\right]$, then by strong duality, $X$ must be $X^{\text{opt}}$ (and furthermore, $Y$ is in fact $Y^{\text{opt}}$, the optimal solution to the dual problem).

A useful criteria for showing that an SDP does satisfy strong duality is given by Slater's theorem:

**Theorem 1** (Slater's theorem for SDPs)**.** *For every SDP,* $\{A, B, \Phi\}$*, the following implications hold:*

1. *If there exists a feasible solution to the primal problem and a* $Y \in \mathcal{B}$ *such that* $\Phi^*\left(Y\right) > A$*, then* $\alpha^{\text{opt}} = \beta^{\text{opt}}$ *and there exists an* $X^{\text{opt}} \in \mathcal{A}$ *such that* $\alpha^{\text{opt}} = \text{Tr}\left[AX^{\text{opt}}\right]$*.*

2. *If there exists a feasible solution to the dual problem and an* $X \in \mathcal{A}$ *such that* $X > 0$*, then* $\alpha^{\text{opt}} = \beta^{\text{opt}}$ *and there exists a* $Y^{\text{opt}} \in \mathcal{B}$ *such that* $\beta^{\text{opt}} = \text{Tr}\left[BY^{\text{opt}}\right]$*.*

*Proof.* For a proof of this theorem, see, for example, [166]. □

**Complementary slackness**

One final property of SDPs that we shall use is that of complementary slackness. This relates the saturation of the dual problem's constraint to the optimal solutions of the primal and dual SDPs.

**Proposition 2** (Complementary slackness for SDPs)**.** *Let* $\{A, B, \Phi\}$ *be a semidefinite program and suppose that* $X \in \mathcal{A}$ *and* $Y \in \mathcal{B}$ *are such that* $\mathrm{Tr}\left[AX\right] = \mathrm{Tr}\left[BY\right]$*. Then:*

$$\Phi^*\left(Y\right)X = AX. \tag{1.9}$$

*Proof.* We have that:

$$\mathrm{Tr}\left[AX\right] = \mathrm{Tr}\left[BY\right] = \mathrm{Tr}\left[\Phi\left(X\right)Y\right] = \mathrm{Tr}\left[\Phi^*\left(Y\right)X\right], \tag{1.10}$$

and hence:

$$\mathrm{Tr}\left[\left(\Phi^*\left(Y\right) - A\right)X\right] = 0. \tag{1.11}$$

As $\Phi^*\left(Y\right) - A$ and $X$ are both positive semidefinite operators and the trace of two positive semidefinite operators is zero if and only if their product is zero,[2] we conclude that $\left(\Phi^*\left(Y\right) - A\right)X = 0$ and we obtain the desired result. □

## 1.3 State discrimination as an SDP

Having introduced semidefinite programming, we now show how it can be applied to minimum error state discrimination. Given a set of states $\mathcal{P} = \{\rho_i\}_{i=1}^k$, each of dimension $d$, prepared according to some probability distribution $\{p_i\}_{i=1}^k$, the goal is to perform the following optimization:

$$\underset{\mathcal{M}=\{M_i\}_{i=1}^k}{\text{Maximize:}} \sum_{i=1}^k \mathrm{Tr}\left[\tilde{\rho}_i M_i\right], \tag{1.12}$$

[2]To see this, let $R$ and $S$ be two positive semidefinite operators. Then $R = TT^\dagger$ and $S = VV^\dagger$ for some positive semidefinite $T$ and $V$. Then:

$$\mathrm{Tr}\left[RS\right] = \mathrm{Tr}\left[TT^\dagger VV^\dagger\right] = \mathrm{Tr}\left[T^\dagger VV^\dagger T\right] \geq 0,$$

where for the last inequality, we have used the fact that $T^\dagger VV^\dagger T$ is a positive semidefinite matrix. Equality holds if and only if $T^\dagger VV^\dagger T = 0$ and hence if and only if $RS = 0$.

where for brevity of notation, we define $\tilde{\rho}_i = p_i \rho_i$. Eq. (1.12) defines the objective function for the primal problem of the state discrimination SDP. The constraints come from the requirement that $\mathcal{M}$ be a valid measurement. For this we require:

$$M_i \geq 0, \quad \forall i, \tag{1.13}$$

$$\sum_{i=1}^{k} M_i = \mathbb{I}. \tag{1.14}$$

We will now show how this can readily be recast in the form of Eq. (1.2) and derive the related dual problem.

### 1.3.1 Formulation as an SDP

Comparing Eqs. (1.12-1.14) with Eq. (1.2), we see that:

- $A$ is a $kd$ by $kd$ block-diagonal matrix with each $d$ by $d$ block, labeled by $i$, given by $\tilde{\rho}_i$:

$$A = \begin{pmatrix} \tilde{\rho}_1 & & \\ & \ddots & \\ & & \tilde{\rho}_k \end{pmatrix}. \tag{1.15}$$

- $B$ is the $d$ by $d$ identity matrix.

- $X$, the variable matrix, is a $kd$ by $kd$ block-diagonal matrix where we label each $d$ by $d$ block by $M_i$:

$$X = \begin{pmatrix} M_1 & & \\ & \ddots & \\ & & M_k \end{pmatrix}. \tag{1.16}$$

- $Y$ is a $d$ by $d$ matrix that we shall call $N$.

- The map $\Phi$ is given by:

$$\Phi(X) = \sum_{i=1}^{k} M_i. \tag{1.17}$$

Using Eq. (1.4), we see that $\Phi^*$ must satisfy:

$$\mathrm{Tr}\left[N \sum_{i=1}^{k} M_i\right] = \mathrm{Tr}\left[\begin{pmatrix} M_1 & & \\ & \ddots & \\ & & M_k \end{pmatrix} \Phi^*(N)\right], \tag{1.18}$$

and hence $\Phi^*(N)$ produces a $kd$ by $kd$ block-diagonal matrix with $N$ in each of the blocks:

$$\Phi^*(N) = \begin{pmatrix} N & & \\ & \ddots & \\ & & N \end{pmatrix}. \tag{1.19}$$

We now have the required information to state the primal and dual SDPs associated with minimum error state discrimination. The primal problem is:

$$\begin{aligned} \underset{\mathcal{M}=\{M_i\}_{i=1}^k}{\text{Maximize:}}\ & \alpha = \sum_{i=1}^{k} \text{Tr}\left[\tilde{\rho}_i M_i\right]. \\ \text{Subject to:}\ & \sum_{i=1}^{k} M_i = \mathbb{I}, \\ & M_i \geq 0, \quad \forall i. \end{aligned} \tag{1.20}$$

The dual is given by:

$$\begin{aligned} \underset{N}{\text{Minimize:}}\ & \beta = \text{Tr}\left[N\right]. \\ \text{Subject to:}\ & N \geq \tilde{\rho}_i, \quad \forall i, \\ & N \in \mathit{Herm}. \end{aligned} \tag{1.21}$$

With these in place, we will now see how strong and weak duality can be used to derive statements about the state discrimination problem.

### 1.3.2 Optimal measurements

Strong duality leads to the following necessary and sufficient conditions for a measurement, $\mathcal{M}$, to be optimal for minimum error state discrimination as originally obtained in [84, 85, 174] and using the SDP formalism in [98, 63]. For those sets of states for which the optimum distinguishing measurement strategy is known, these conditions have often been used to derive it.

**Theorem 2.** *Suppose an unknown state, $\sigma$, is prepared from a set of known states, $\mathcal{P} = \{\rho_i\}_{i=1}^k$, according to a probability distribution, $\{p_i\}_{i=1}^k$. A measurement, $\mathcal{M} = \{M_i\}_{i=1}^k$, is optimal for attempting to identify the preparation with minimum error if and only if:*

$$N = \sum_{i=1}^{k} \tilde{\rho}_i M_i, \tag{1.22}$$

*is Hermitian and satisfies $N \geq \tilde{\rho}_i$, for all $i$.*

*Proof.* First note that the state discrimination SDP as defined in Eqs. (1.20) and (1.21) satisfies the conditions of Slater's theorem. To see this, consider $M_i = \frac{1}{k}\mathbb{I}$ for all $i$ and $N = 2\mathbb{I}$. Each $M_i$ is strictly positive definite and so they strictly satisfy the constrains of Eq. (1.20). Furthermore, it is clear that such an $N$ is Hermitian and as, for each $i$, the eigenvalues of $\tilde{\rho}_i$ are certainly less than or equal to 1, we have $N > \tilde{\rho}_i$ for all $i$. Hence, $N$ strictly satisfies the constraints of the dual problem. As the conditions of Slater's theorem are satisfied, we know that for this SDP strong duality holds and hence that $\alpha^{\text{opt}}=\beta^{\text{opt}}$ and there exist feasible choices of $\{M_i\}_{i=1}^k$ and $N$ that achieve these values.

Suppose we are given a valid measurement, $\mathcal{M} = \{M_i\}_{i=1}^k$, and that the $N$ defined by Eq. (1.22) satisfies the constraints of the dual problem. Then:

$$\beta = \text{Tr}\left[N\right] = \text{Tr}\left[\sum_{i=1}^{k} \tilde{\rho}_i M_i\right] = \sum_{i=1}^{k} \text{Tr}\left[\tilde{\rho}_i M_i\right] = \alpha.$$

Hence, by strong duality, $\mathcal{M}$ is an optimal measurement.

Now suppose $\mathcal{M}$ is an optimal measurement. By complementary slackness (Proposition 2), if $N$ is an optimal solution to the dual problem, it satisfies:

$$\Phi^*\left(N\right)\begin{pmatrix} M_1 & & \\ & \ddots & \\ & & M_k \end{pmatrix} = \begin{pmatrix} \tilde{\rho}_1 M_1 & & \\ & \ddots & \\ & & \tilde{\rho}_k M_k \end{pmatrix},$$

$$\Rightarrow \quad \begin{pmatrix} NM_1 & & \\ & \ddots & \\ & & NM_k \end{pmatrix} = \begin{pmatrix} \tilde{\rho}_1 M_1 & & \\ & \ddots & \\ & & \tilde{\rho}_k M_k \end{pmatrix},$$

which implies that:

$$NM_i = \tilde{\rho}_i M_i, \quad \forall i.$$

Taking the sum over $i$ on both sides and using the fact that $\sum_{i=1}^k M_i = \mathbb{I}$, we obtain:

$$N = \sum_{i=1}^{k} \tilde{\rho}_i M_i.$$

□

### 1.3.3 Bounds on the probability of success

For the vast majority of state sets, the optimal discrimination measurement is not known. In these cases, bounds on the success probability can be obtained. Weak duality enables us to

reproduce some upper bounds on the success probability, $p_{\text{succ}}$, using the SDP framework. In particular:

**Lemma 1.** *Suppose an unknown state, $\sigma$, is prepared from a set of known states, $\mathcal{P} = \{\rho_i\}_{i=1}^k$, according to a probability distribution, $\{p_i\}_{i=1}^k$. The maximum probability of correctly identifying the preparation, $p_{\text{succ}}$, satisfies:*[3]

1. *From [83]:*

$$p_{\text{succ}} \leq d \max_i \|\tilde{\rho}_i\| . \tag{1.23}$$

2. *From [143]:*

$$p_{\text{succ}} \leq \frac{1}{2} + \frac{1}{2} \frac{1}{k-1} \sum_{1 \leq i < j \leq k} \text{Tr} \left| \tilde{\rho}_i - \tilde{\rho}_j \right| . \tag{1.24}$$

3. *From [144]:*

$$p_{\text{succ}} \leq \min_i \left\{ p_i + \sum_{j \neq i} \text{Tr} \left[ \left( \tilde{\rho}_j - \tilde{\rho}_i \right)_+ \right] \right\} . \tag{1.25}$$

4. *From [135]:*

$$p_{\text{succ}} \leq \text{Tr} \left[ \sqrt{\sum_{i=1}^k \tilde{\rho}_i^2} \right] . \tag{1.26}$$

5. *From [9]:*

   *Given two Hermitian operators, $A$ and $B$, define:*

$$\max(A, B) = \frac{1}{2} \left[ A + B + |A - B| \right] . \tag{1.27}$$

   *For a permutation, $\epsilon$, acting on $k$ objects, taken from the permutation group $S_k$, consider:*

$$N_\epsilon = \max \left( \tilde{\rho}_{\epsilon(k)}, \max \left( \tilde{\rho}_{\epsilon(k-1)}, \max \left( \ldots, \max \left( \tilde{\rho}_{\epsilon(2)}, \tilde{\rho}_{\epsilon(1)} \right) \right) \right) \right) . \tag{1.28}$$

   *Then:*

$$p_{\text{succ}} \leq \min_{\epsilon \in S_k} \text{Tr} \left[ N_\epsilon \right] . \tag{1.29}$$

[3] In what follows:

- $\|A\|$ is the operator norm, the square root of the largest eigenvalue of $A^\dagger A$.
- $|A| = \sqrt{A^\dagger A}$.
- Suppose $A$ has spectral decomposition $A = \sum_{i=1}^d \lambda_i |u_i\rangle\langle u_i|$. Then $A_+ = \sum_{i:\lambda_i > 0} \lambda_i |u_i\rangle\langle u_i|$.

*Proof.* The general aim is to construct an $N$ that satisfies the constraints of the dual problem given in Eq. (1.21). Then, by weak duality, we have that $p_{\text{succ}} \leq \text{Tr}\,[N]$.

1. Let $\lambda^{\max} = \max_i \|\tilde{\rho}_i\|$. To obtain the bound, define:

$$N = \lambda^{\max} \mathbb{I}_d.$$

Note that taking the trace of $N$ gives Eq. (1.23). To see that $N \geq \tilde{\rho}_i$, consider:

$$\begin{aligned} N - \tilde{\rho}_i &= \sum_{j=1}^{d} \left(\lambda^{\max} - \lambda_j^i\right) |u_j^i\rangle\langle u_j^i|, \\ &\geq 0, \end{aligned}$$

where $\sum_{j=1}^{d} \lambda_j^i |u_j^i\rangle\langle u_j^i|$ is the spectral decomposition of $\tilde{\rho}_i$ and we have written $\mathbb{I}_d = \sum_{j=1}^{d} |u_j^i\rangle\langle u_j^i|$.

2. Consider $\max(A, B)$ as defined in Eq. (1.27). Note that $\max(A, B) \geq A$ and $\max(A, B) \geq B$ as:

$$\begin{aligned} \max(A, B) - A &= \frac{1}{2}\left[B - A + |A - B|\right], \\ &= \frac{1}{2}\left[\sum_{i=1}^{d} \lambda_i |u_i\rangle\langle u_i| + \sum_{i=1}^{d} |\lambda_i|\, |u_i\rangle\langle u_i|\right], \\ &\geq 0, \end{aligned}$$

where $\sum_{i=1}^{d} \lambda_i |u_i\rangle\langle u_i|$ is the spectral decomposition of $B - A$.

To obtain the bound, define:

$$\begin{aligned} N &= \frac{1}{k-1} \sum_{1 \leq i < j \leq k} \max(\tilde{\rho}_i, \tilde{\rho}_j), \\ &= \frac{1}{2} \sum_{i=1}^{k} \tilde{\rho}_i + \frac{1}{2} \frac{1}{k-1} \sum_{1 \leq i < j \leq k} |\tilde{\rho}_i - \tilde{\rho}_j|. \end{aligned}$$

Note that taking the trace of $N$ gives Eq. (1.24). To see that $N \geq \tilde{\rho}_i$, consider, without loss of generality:

$$\begin{aligned} N - \tilde{\rho}_1 &= \frac{1}{k-1} \sum_{j=2}^{k} \left[\max(\tilde{\rho}_1, \tilde{\rho}_j) - \tilde{\rho}_1\right] + \sum_{2 \leq i < j \leq k} \max(\tilde{\rho}_i, \tilde{\rho}_j), \\ &\geq 0. \end{aligned}$$

3. To obtain the bound, define:

$$N_i = \tilde{\rho}_i + \sum_{j \neq i} \left(\tilde{\rho}_j - \tilde{\rho}_i\right)_+ ,$$

and note that taking the trace of $N_i$ and minimizing over $i$ gives Eq. (1.25). To see that $N_i \geq \tilde{\rho}_r$, consider:

$$\begin{aligned} N_i - \tilde{\rho}_r &= \tilde{\rho}_i - \tilde{\rho}_r + \sum_{j \neq i} \left(\tilde{\rho}_j - \tilde{\rho}_i\right)_+ , \\ &= \left(\tilde{\rho}_i - \tilde{\rho}_r\right) + \left(\tilde{\rho}_r - \tilde{\rho}_i\right)_+ + \sum_{j \neq i,r} \left(\tilde{\rho}_j - \tilde{\rho}_i\right)_+ , \\ &= \sum_{s=1}^{d} \lambda_i |u_s\rangle\langle u_s| + \sum_{s:\lambda_s<0} |\lambda_s| \, |u_s\rangle\langle u_s| + \sum_{j \neq i,r} \left(\tilde{\rho}_j - \tilde{\rho}_i\right)_+ , \\ &\geq 0, \end{aligned}$$

where $\sum_{s=1}^{d} \lambda_s |u_s\rangle\langle u_s|$ is the spectral decomposition of $\tilde{\rho}_i - \tilde{\rho}_r$.

4. To obtain the bound, define:

$$N = \sqrt{\sum_{i=1}^{k} \tilde{\rho}_i^2},$$

and note that taking the trace of $N$ gives Eq. (1.26). To see that $N \geq \tilde{\rho}_i$, consider:

$$\sum_{j=1}^{k} \tilde{\rho}_j^2 \geq \tilde{\rho}_i^2.$$

The square root function is an operator monotone and hence:

$$\sqrt{\sum_{j=1}^{k} \tilde{\rho}_j^2} \geq \tilde{\rho}_i,$$

as required.

5. To obtain the bound, construct $N_\epsilon$ iteratively as follows:

$$\begin{aligned} N_2 &= \max\left(\tilde{\rho}_{\epsilon(2)}, \tilde{\rho}_{\epsilon(1)}\right) \\ N_3 &= \max\left(\tilde{\rho}_{\epsilon(3)}, N_2\right) \\ \vdots &= \vdots \\ N_\epsilon = N_k &= \max\left(\tilde{\rho}_{\epsilon(k)}, N_{k-1}\right). \end{aligned}$$

Using the fact that $\max(A, B) \geq A$ and $\max(A, B) \geq B$, by construction we have $N_\epsilon \geq \tilde{\rho}_i$. By minimizing over all $\epsilon \in S_k$, we obtain Eq. (1.29).

□

## 1.4 Summary

In this chapter we have introduced the task of minimum error state discrimination and shown how it can formulated using the tool of semidefinite programming. As well as providing an efficient numerical solution for any given state discrimination problem, we have illustrated how the semidefinite programming approach can be used to derive tests for the optimality of a measurement and bounds on how well it is possible to distinguish between a set of states. Having access to such tests and bounds is important for analyzing the strengths and weaknesses of communication and cryptographic protocols.

A more fundamental question regarding state discrimination is to ask why perfect discrimination is impossible for the vast majority of sets of states? What is it about quantum mechanics that enables us to achieve only a certain level of success? Is there some underlying mechanism that provides a more intuitive explanation for this? A potential avenue for providing answers to this is that of $\psi$-epistemic models of quantum theory and we turn to these in the next chapter.

# Chapter 2

# State Exclusion

## 2.1 An explanation for indistinguishability

What is the quantum state, $|\psi\rangle$? Attempting to answer this question has led to numerous different interpretations of quantum mechanics, each with their pros and cons.

The framework of ontological models [82] aims to pose the above questions in a more rigorous manner. Positing that when we prepare a system in a quantum state, there exists an underlying physical state, $\lambda$, providing a complete description of the system, ontological models provide a tool to analyze the relationship between such a $\lambda$ and $|\psi\rangle$. Three potential options arise:

1. $|\psi\rangle$ is in one-to-one correspondence with $\lambda$. The quantum state *is* the complete description of the system.

2. Many $\lambda$ are associated with a given $|\psi\rangle$ and each $\lambda$ is associated with a unique $|\psi\rangle$. In essence, $\lambda$ can be decomposed as $\lambda = (|\psi\rangle, \omega)$ where $\omega$ is a hidden variable needed in addition to $|\psi\rangle$ to completely specify the state of the system.

3. Many $\lambda$ are associated with a given $|\psi\rangle$ and each $\lambda$ can potentially be associated with many different $|\psi\rangle$.

Models that fall into categories 1 or 2 are said to be *$\psi$-ontic*. Here the quantum state is a physical property of the system in the sense that if one had access to the description of the system given by $\lambda$, one would be able to deduce the system's quantum state. Those models in the third category are referred to as *$\psi$-epistemic* and the quantum state can be regarded as 'merely' capturing our uncertainty about the true physical state of the system: it is a state of

knowledge rather than a physical property.

A $\psi$-epistemic ontological model of quantum mechanics has the potential to alleviate many of the paradoxes of quantum theory and perhaps provide insight into the origin of some of its distinctive attributes. For example, the instantaneous collapse of the quantum state under measurement, and the precise mechanism and timing surrounding it, is problematic and poorly understood if the quantum state is taken to be a physical object or property. If however, it is regarded as a state of knowledge of the physical system, existing only in the mind of an observer, then its updating upon them learning the outcome of a measurement is almost to be expected.

Similarly, a $\psi$-epistemic model may give insight into why non-orthogonal quantum pure states cannot be perfectly distinguished from a single-shot measurement. Even if we somehow had knowledge of the complete physical description of the system, in a $\psi$-epistemic model we would still not be able to identify the quantum state of the system with certainty. Perhaps this can provide the answer to the question posed at the end of the previous chapter? An epistemic toy model by Spekkens [159], that reproduces a stripped-down version of quantum theory, suggests that this may indeed be the case.[1]

However, a recent no-go theorem by Pusey, Barrett and Rudolph (PBR) [142] shows that, subject to certain assumptions on how the physical state of independently prepared systems behave, a $\psi$-epistemic model of quantum theory is impossible, somewhat denting the explanatory power of such a construction.

In the rest of this section, we will introduce the formalism of ontological models and sketch how Spekkens' toy bit hints at the possibility of $\psi$-epistemic models providing an explanation for the impossibility of perfect quantum state discrimination. We will then examine the PBR argument and this will lead us into the next section where we discuss the task of quantum state exclusion.

### 2.1.1 Ontological models

Ontological models provide a formalism for classifying hidden variable theories within an operational framework and were originally introduced in [158, 82]. They suppose that when a system is prepared in a quantum state, $|\psi\rangle$, the system is actually in some physical state, $\lambda$,

[1] In addition to the impossibility of state discrimination, the model in [159] exhibits other quantum properties such as entanglement, teleportation and dense coding.

that potentially gives a more complete description of the system than $|\psi\rangle$. This physical state is often referred to as the *ontic* state of the system. The set of all ontic states is denoted by $\Lambda$.

A given quantum preparation could potentially result in a number of different ontic states with the probability of each denoted by $\mu_\psi(\lambda)$. In this case, we can interpret $|\psi\rangle$ as capturing our partial information about the underlying physical state and $\mu_\psi(\lambda)$ is referred to as the *epistemic* state of the system. The epistemic state satisfies:

$$\begin{aligned} \mu_\psi(\lambda) &\geq 0, \quad \forall |\psi\rangle,\ \forall \lambda \in \Lambda, \\ \int_{\lambda\in\Lambda} \mu_\psi(\lambda)\, d\lambda &= 1, \quad \forall |\psi\rangle, \end{aligned} \tag{2.1}$$

ensuring that $\mu_\psi(\lambda)$ defines a valid probability distribution. Though we shall not need it in the following discussion, note that the above definition can be extended to cover mixed quantum states, $\rho$. For such states however, the epistemic state need not be uniquely determined by $\rho$ and the precise decomposition of $\rho$ into pure states may be required. Models with this feature are termed *preparation contextual* [158].

To characterize what happens when a measurement is performed on the system within an ontological model, we require that the probabilities of the individual measurement results are determined solely by the ontic state the system was prepared in. If a measurement, $\mathcal{M} = \{M_i\}_{i=1}^k$, has possible outcomes $i \in \{1, \ldots, k\}$, then the probability of obtaining outcome $j$ when the measurement $\mathcal{M}$ is performed on a system in an ontic state $\lambda$ is given by a *response function*, $\xi_\mathcal{M}(j|\lambda)$. These response functions satisfy:

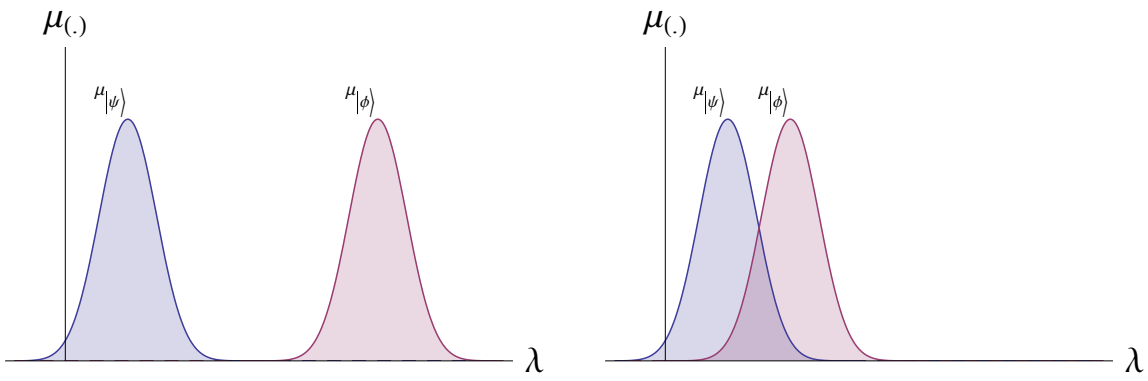
$$\begin{aligned} \xi_\mathcal{M}(i|\lambda) &\geq 0, \quad \forall \mathcal{M},\ \forall i,\ \forall \lambda \in \Lambda, \\ \sum_{i=1}^{k} \xi_\mathcal{M}(i|\lambda) &= 1, \quad \forall \mathcal{M},\ \forall \lambda \in \Lambda. \end{aligned} \tag{2.2}$$

This ensures that the probability of the measurement outcomes forms a valid probability distribution.

Eqs. (2.1) and (2.2) define constraints on the relationship between an ontological model and preparations and measurements in quantum theory. Notions of dynamics and transformations can also be defined though will not be needed here [158, 110]. For a model to reproduce the measurement predictions of quantum mechanics, the following must hold:

$$\int_{\lambda\in\Lambda} \mu_\psi(\lambda)\, \xi_\mathcal{M}(i|\lambda)\, d\lambda = \langle\psi|M_i|\psi\rangle, \quad \forall |\psi\rangle,\ \forall \mathcal{M},\ \forall i. \tag{2.3}$$

As introduced previously, there are two distinct classes of ontological models: $\psi$-ontic and $\psi$-epistemic. Let $\Lambda_\psi = \{\lambda : \mu_\psi(\lambda) > 0\}$ denote the set of all ontic states that could be produced




(a) $\psi$-ontic model of $|\psi\rangle$ and $|\phi\rangle$.

(b) $\psi$-epistemic model of $|\psi\rangle$ and $|\phi\rangle$.

Figure 2.1: *Classes of ontological models.* Relationships between $\mu_\psi(\lambda)$ and $\mu_\phi(\lambda)$ for different classes of ontological models.

when the state $|\psi\rangle$ is prepared. If for all pairs of distinct quantum pure states, $|\psi\rangle$ and $|\phi\rangle$, it holds that $\Lambda_\psi \cap \Lambda_\phi = \emptyset$, then the model is $\psi$-ontic.[2] This is illustrated in Figure 2.1a and here the quantum state can be uniquely determined if one has knowledge of the ontic state of the system.

For a model to be classified as $\psi$-epistemic, there must exist at least one pair of distinct $|\psi\rangle$ and $|\phi\rangle$ such that $\Lambda_\psi \cap \Lambda_\phi \neq \emptyset$. This is illustrated in Figure 2.1b and here, even given knowledge of the ontic state of the system, it can be impossible to determine which quantum state the system was prepared in.

We now move onto consider a toy ontological model that gives a suggestive hint towards the potential explanatory power of an epistemic model.

### 2.1.2 Spekkens' toy bit and indistinguishability

In [159], a toy theory is introduced that reproduces the predictions of quantum mechanics for qubits which can only be prepared and measured in one of the $x$, $y$ and $z$ bases. The underlying ontic state space for this model consists of four states which, following the exposition of [110], we denote by $\lambda_1 = (-,-), \lambda_2 = (-,+), \lambda_3 = (+,-)$ and $\lambda_4 = (+,+)$.

To construct the epistemic states of the theory, Spekkens imposes a constraint on the allowed pure state preparations. This is termed the *knowledge balance principle* and effectively imposes

[2] More formally, one can define a $\psi$-ontic model to be an ontological model such that for each pair $|\psi\rangle$ and $|\phi\rangle$, it holds that $\Lambda_\psi$ and $\Lambda_\phi$ overlap on a set of zero measure.

that while the ontic state contains the most complete description of the system, we will only ever be able to determine that the system is in one of two equally likely ontic states. Hence, there are six pure epistemic states that one can prepare, one for each pair of ontic states. We denote them by $|x\pm)$, $|y\pm)$ and $|z\pm)$ and illustrate them in Figure 2.2. Note that the theory is '$\psi$'-epistemic as the epistemic states for distinct pure state preparations may overlap.

Measurements within the toy theory are assumed to be repeatable; if we perform the same measurement twice in succession, we should obtain the same outcome. They must also respect the knowledge balance principle so that after a measurement of a pure state, the epistemic state of the system again consists of two ontic states. These requirements impose that there are only three, 2-outcome, measurements that can be performed on the system, $X, Y$ and $Z$. Labeling the outcomes by $\pm 1$, the measurement probabilities satisfy:

$$P\left(s = \pm 1 || r\pm)\right) = \begin{cases} \delta_{\pm,\pm} & \text{, if } r = s, \\ \frac{1}{2} & \text{, if } r \neq s, \end{cases} \tag{2.4}$$

for $r, s \in \{x, y, z\}$.

These measurement statistics reproduce the indistinguishability of quantum theory. For example, given a system prepared in either $|x+)$ or $|y-)$, none of the three allowable measurements enable us to determine which was prepared. In both cases, there is 50% chance that the system was prepared in the ontic state $(+, -)$, explaining why we cannot detect the difference between them. Furthermore, these states are analogous to the quantum states $|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right)$ and $|i-\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - i|1\rangle\right)$, so perhaps the $\psi$-epistemic nature of the model induced by the knowledge balance principle can be extended to explain the indistinguishability inherent in quantum mechanics?

However, in the next section, we shall see that, subject to some very plausible assumptions, a $\psi$-epistemic model of quantum theory is impossible.

### 2.1.3 The PBR argument

While a $\psi$-epistemic model underlying quantum theory would potentially provide an intuition for many of the strange features of quantum mechanics, in recent years a number of results have shown, subject to varying assumptions, that an ontological model of quantum theory must be $\psi$-ontic [46, 142, 47, 81, 1, 48]. In particular, the result of Pusey, Barrett and Rudolph [142] kick-started the debate and we review it here.

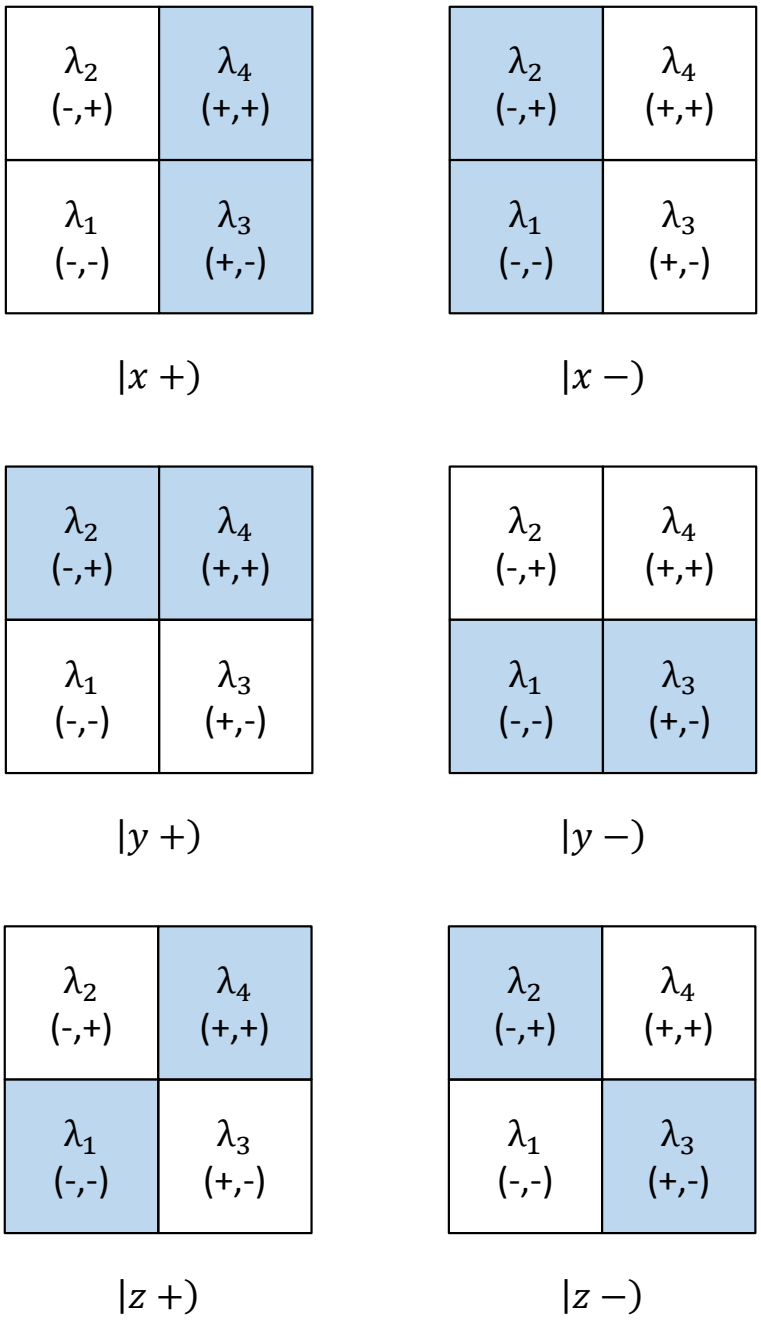



Figure 2.2: *The epistemic states of Spekkens' toy bit.* There are six epistemic states for Spekkens' toy bit. In each instance, the ontic states shaded in blue belong to $\Lambda_{r\pm}$.

Results indicating the impossibility of $\psi$-epistemic theories aim to show that, for such models, Eq. (2.3) cannot be satisfied and hence the complete set of measurement predictions of quantum mechanics cannot be reproduced. To show this, PBR require one additional assumption, referred to in [110] as the *preparation independence postulate*. This asserts that if two systems are prepared in quantum states independently of one another, then the ontic states of the systems should also be uncorrelated. So, if $n$ systems are each independently prepared in state $|\psi_i\rangle$ for $1 \leq i \leq n$, the global quantum state is:

$$|\Psi\rangle = \bigotimes_{i=1}^{n} |\psi_i\rangle, \tag{2.5}$$

then the global ontic state is:

$$\bar{\lambda} = (\lambda_1, \ldots, \lambda_n), \tag{2.6}$$

where $\lambda_i$ is the ontic state resulting from the quantum preparation on system $i$ and is independent of each other preparation. Under this assumption, for quantum product states, the ontic state of the global system is specified by listing the ontic states of the individual systems and the global epistemic state is given by $\mu_{|\Psi\rangle}\left(\bar{\lambda}\right) = \mu_{|\psi_1\rangle}(\lambda_1) \ldots \mu_{|\psi_n\rangle}(\lambda_n)$.

With this in place, the PBR argument then runs as follows: consider $n$ independent copies of a device that prepares one of two states, $|\psi_0\rangle$ or $|\psi_1\rangle$, each with probability 1/2. Without loss of generality, these can be regarded as being separated by an angle $\theta$, defined on qubits and taken to be:

$$\begin{aligned} |\psi_0(\theta)\rangle &= \cos\left(\frac{\theta}{2}\right)|0\rangle + \sin\left(\frac{\theta}{2}\right)|1\rangle, \\ |\psi_1(\theta)\rangle &= \cos\left(\frac{\theta}{2}\right)|0\rangle - \sin\left(\frac{\theta}{2}\right)|1\rangle, \end{aligned} \tag{2.7}$$

where $0 \leq \theta \leq \pi/2$. The global quantum state of the $n$ systems is then:

$$|\Psi_{\vec{x}}(\theta)\rangle = \bigotimes_{i=1}^{n} |\psi_{x_i}(\theta)\rangle, \tag{2.8}$$

where $\vec{x} \in \{0,1\}^n$ and each of the $2^n$ possible preparations is equally likely.

By definition, in a $\psi$-epistemic theory, for at least one value of $\theta$, it must hold that $\Lambda_{\psi_0} \cap \Lambda_{\psi_1} \neq \emptyset$. For some preparations, even if we knew the ontic state the system had been prepared in, we would not be able to deduce with certainty whether the device had produced $|\psi_0\rangle$ or $|\psi_1\rangle$. Suppose such a preparation happens with probability $q$ so that with probability $q^n$, it will occur for each of the $n$ systems under consideration. In this situation, knowing the ontic state of each

of the $n$ systems would not only give us insufficient information to identify the global quantum state $|\Psi_{\vec{x}}\rangle$ but we would also not have enough information to identify a single preparation that *had not taken place*. According to our information, any one of the $2^n$ preparations could have occurred.

However, for every value of $\theta$, PBR show that there exists an $n$ such that we can find a quantum measurement, $\mathcal{M} = \{|\xi_{\vec{x}}\rangle\langle\xi_{\vec{x}}|\}_{\vec{x}\in\{0,1\}^n}$, such that the probability of obtaining the outcome labeled $\vec{y}$ when the global quantum state of the system is $|\Psi_{\vec{y}}\rangle$, is zero, for all $\vec{y}$. In other words:

$$\left|\langle\xi_{\vec{y}}|\Psi_{\vec{y}}\rangle\right|^2 = 0, \quad \forall\vec{y}. \tag{2.9}$$

According to a $\psi$-epistemic model, with probability $q^n$ we should regard all preparations as possible. However, $\mathcal{M}$ is such that when we see outcome $\vec{y}$, we know with certainty that the preparation was not $|\Psi_{\vec{y}}\rangle$. Hence, there is a contradiction between the predictions of quantum mechanics encapsulated in Eq. (2.9) and those of $\psi$-epistemic models satisfying the preparation independence postulate.

More precisely, PBR show that quantum mechanics admits such a measurement for a given $\theta$, provided that $n$ is chosen large enough for:

$$2^{1/n} - 1 \leq \tan\left(\frac{\theta}{2}\right), \tag{2.10}$$

to hold. This raises the question as to whether $n$ can be taken to be any smaller in deriving the PBR result? To answer this question, we shall develop the problem of conclusive state exclusion, investigating when it is possible to find a measurement for a given set of states such that a relation similar to Eq. (2.9) holds. The rest of this chapter shall be dedicated towards this goal.

Furthermore, other no-go theorems from the foundations of quantum mechanics, such as Bell's theorem, have led to striking information theoretic consequences showing separations between the power of quantum and classical resources. Can similar results be derived from the PBR theorem? This will be the concern of Chapter 3.

### 2.1.4 Aside: The impossibility of maximally epistemic models

Note that the PBR result (and other no-go theorems against $\psi$-epistemic models) require additional assumptions about the structure and dynamics of the ontic state space. Indeed, it is in fact possible to construct a $\psi$-epistemic model of quantum theory [111, 1]. Naturally, these

models violate the underlying assumptions of the no-go theorems such as the preparation independence postulate. Can such constructions fully explain the indistinguishability of quantum states?

To answer this question, following [17], we need to define what it means for a $\psi$-epistemic model to completely explain quantum indistinguishability. More precisely, for quantum states $|\psi\rangle$ and $|\phi\rangle$, the overlap of the epistemic states, $\omega_C$, is quantified by the classical overlap of the probability distributions $\mu_\psi(\lambda)$ and $\mu_\phi(\lambda)$:

$$\omega_C(\mu_\psi, \mu_\phi) = 1 - \frac{1}{2}\int |\mu_\psi(\lambda) - \mu_\phi(\lambda)|\, d\lambda. \tag{2.11}$$

This quantity has an operational interpretation. Suppose $|\psi\rangle$ and $|\phi\rangle$ were each prepared with probability 1/2. Then given knowledge of the ontic state $\lambda$, the probability of correctly identifying which preparation took place is given by $1 - \frac{1}{2}\omega_C(\mu_\psi, \mu_\phi)$.

Similarly, the overlap between the quantum states is denoted by $\omega_Q(\psi, \phi)$ and given by:

$$\omega_Q(\psi, \phi) = 1 - \sqrt{1 - |\langle\psi|\phi\rangle|^2}. \tag{2.12}$$

If $|\psi\rangle$ and $|\phi\rangle$ are prepared with equal probability, the probability of correctly identifying the preparation by performing the optimal quantum measurement on the system, is given by $1 - \frac{1}{2}\omega_Q(\psi, \phi)$.

An ontological model of quantum theory is said to be maximally $\psi$-epistemic if for all pairs of states:

$$\omega_C(\mu_\psi, \mu_\phi) = \omega_Q(\psi, \phi). \tag{2.13}$$

This would give an explanation for the indistinguishability inherent in quantum mechanics as even if we had access to the more complete description provided by $\lambda$, we would not improve our chance of identifying a given preparation correctly. However, it was shown in [119, 17], without additional assumptions such as preparation independence, that for systems of dimension larger than 2, no ontological model that reproduces quantum theory can be maximally $\psi$-epistemic. Interestingly, the proof also makes use of state exclusion measurements, which we turn to in the next section.

## 2.2 The task of state exclusion

In state discrimination, we effectively attempt to increase our knowledge of the system so that we progress from knowing it is one of $k$ possibilities, to knowing it is one particular state. We

reduce the set of possible preparations that could have occurred from $k$ to 1. A related, and less ambitious, task would be to exclude $m$ preparations from the set, reducing the size of the set of possible states from $k$ to $k-m$. If we rule out the $m$ states with certainty, we say that they have been conclusively excluded.

As discussed in Section 2.1.3, single-state exclusion ($m = 1$) has found use in proving results in the foundations of quantum mechanics [142, 119, 17]. It has also previously been considered with respect to quantum state compatibility criteria [39] and displays similarities with the problem of finding quantum strategies to hedge bets [123, 7]. Its unambiguous variant, which we shall discuss in Section 2.4.1, has been used to construct schemes for quantum digital signatures [50].

More formally, what does it mean to be able to perform conclusive exclusion? We first consider the case of single-state exclusion and then show how it generalizes to $m$-state exclusion. As before when considering discrimination, let the set of possible preparations on a $d$-dimensional quantum system be $\mathcal{P} = \{\rho_i\}_{i=1}^k$ and let each preparation occur with probability $p_i$. Again we define $\tilde{\rho}_i = p_i\rho_i$ and call the prepared state $\sigma$. The aim is to perform a measurement on $\sigma$ so that, from the outcome, we can state a $j \in \{1, \ldots, k\}$ such that $\sigma \neq \rho_j$.

Such a measurement will consist of $k$ measurement operators, one for attempting to exclude each element of $\mathcal{P}$. We want a measurement, described by $\mathcal{M} = \{M_i\}_{i=1}^k$, that never leads us to state $j$ when $\sigma = \rho_j$. We need:

$$\mathrm{Tr}\left[\rho_i M_i\right] = 0, \quad \forall i, \tag{2.14}$$

or equivalently, since $\rho_i$ and $M_i$ are positive semidefinite matrices and $p_i$ is a positive number:

$$\alpha = \sum_{i=1}^{k} \mathrm{Tr}\left[\tilde{\rho}_i M_i\right] = 0. \tag{2.15}$$

There will be some instances of $\mathcal{P}$ for which an $\mathcal{M}$ cannot be found to satisfy Eq. (2.15). In these cases our goal is to minimize $\alpha$, which corresponds to the probability of failure of the strategy, 'if outcome $j$ occurs, say $\sigma \neq \rho_j$'.

To formulate $m$-state exclusion, we proceed as follows. Define $Y_{(k,m)}$ to be the set of all subsets of $[k]$ of size $m$. The aim is to perform a measurement on $\sigma$ such that from the outcome we can state a set, $Y \in Y_{(k,m)}$, such that $\sigma \notin \{\rho_y\}_{y\in Y}$. Such a measurement, denoted $\mathcal{M}_m$, will consist of $\binom{k}{m}$ measurement operators and we require that, for each set $Y$:

$$\mathrm{Tr}\left[\tilde{\rho}_y M_Y\right] = 0, \quad \forall y \in Y. \tag{2.16}$$

If we define:

$$\hat{\rho}_Y = \sum_{y \in Y} \tilde{\rho}_y, \tag{2.17}$$

then this can be reformulated as requiring:

$$\mathrm{Tr}\left[\hat{\rho}_Y M_Y\right] = 0, \quad \forall Y \in Y_{(k,m)}. \tag{2.18}$$

Eq. (2.18) is identical in form to Eq. (2.14). Hence we can view $m$-state exclusion as single-state exclusion on the set $\mathcal{P}_m = \{\hat{\rho}_Y\}_{Y \in Y_{(k,m)}}$. Furthermore, we can generalize this approach to an arbitrary collection of subsets that are not necessarily of the same size. With this in mind, in most of what follows, we will restrict ourselves to considering single state exclusion.

### 2.2.1 Exclusion as discrimination and vice versa

The tasks of state exclusion and state discrimination share many similarities. Indeed, if we were instead trying to maximize $\alpha$ in Eq. (2.15), we would obtain the objective function of the primal SDP for state discrimination in Eq. (1.20). More connections will become apparent when we formulate the SDP for state exclusion in Section 2.3.

It is also possible to recast each problem as an instance of the other. Firstly, state discrimination can be put in the form of an exclusion problem by taking $m = k - 1$. If we can exclude $k - 1$ of the possible states, then we can identify $\sigma$ as the remaining state.

Secondly, following the observation of [129] regarding minimum Bayes cost problems, state exclusion can be converted into a discrimination task. To see this, from $\mathcal{P}$ define:

$$\mathcal{R} = \left\{ \vartheta_i = \frac{1}{k-1} \sum_{j \neq i} \tilde{\rho}_j \right\}_{i=1}^{k}. \tag{2.19}$$

Writing $p^{\mathrm{D}}_{\mathrm{error}}$ and $p^{\mathrm{E}}_{\mathrm{error}}$ to distinguish between the probability of error in discrimination and exclusion respectively, in state discrimination on $\mathcal{R}$ we would attempt to minimize:

$$\begin{aligned}
p^{\mathrm{D}}_{\mathrm{error}}(\mathcal{R}) &= 1 - \sum_{i=1}^{k} \mathrm{Tr}\left[\vartheta_i M_i\right], \\
&= 1 - \sum_{i=1}^{k} \sum_{j \neq i} \frac{1}{k-1} \mathrm{Tr}\left[\tilde{\rho}_j M_i\right], \\
&= 1 - \frac{1}{k-1} \sum_{i=1}^{k} \sum_{j=1}^{k} \mathrm{Tr}\left[\tilde{\rho}_j M_i\right] + \frac{1}{k-1} \sum_{i=1}^{k} \mathrm{Tr}\left[\tilde{\rho}_i M_i\right], \\
&= \frac{k-2}{k-1} + \frac{1}{k-1} p^{\mathrm{E}}_{\mathrm{error}}(\mathcal{P}).
\end{aligned}$$

Hence, minimizing the error probability in discrimination on $\mathcal{R}$ is equivalent to minimizing the probability of error in state exclusion on $\mathcal{P}$, and the optimal measurement is the same for both. This interplay between the two tasks enables us to apply bounds on the error probability of state discrimination, such as those in Lemma 1, to the task of state exclusion.

### 2.2.2 Geometric interpretation

Given a set of pure states, $\mathcal{P} = \{|\psi_i\rangle\}_{i=1}^k$, when can conclusive exclusion be performed by a projective measurement, $\mathcal{M} = \{|i\rangle\langle i|\}_{i=1}^k$? It was shown in [39] that $\mathcal{P}$ must be such that there exists a basis in which:

$$|\psi_i\rangle = \sum_{j\neq i} a_j^{(i)} |i\rangle, \quad \forall i. \tag{2.20}$$

This obviously satisfies $\langle i|\psi_i\rangle = 0$, for all $i$, and hence conclusive exclusion is possible. Figure 2.3 illustrates this result for $d = k = 3$ and when all of the coefficients $a_j^{(i)}$ are real.

## 2.3 State exclusion as an SDP

As recognized in [142] for the case of single-state exclusion, the problem of conclusive exclusion can be formulated as an SDP. We saw in Section 1.3 that this formalism was useful in deriving many results regarding the task of minimum error state discrimination. As this forms a subclass of the general exclusion framework, it is reasonable to expect that a similar approach will pay dividends here.

In this section we will use semidefinite programming to produce necessary and sufficient conditions for an exclusion measurement to be optimal. This is analogous to Theorem 2 for state discrimination. By applying these requirements to the PBR exclusion problem, we will be able to determine whether Eq. (2.10) is necessary as well as sufficient. As for state discrimination, weak duality will enable us to derive bounds on the probability of success in an exclusion task and on top of this, we will obtain a necessary condition a set of states must satisfy for conclusive exclusion to be possible.

To obtain the optimal measurement strategy for single-state exclusion, our goal is to minimize $\alpha$ in Eq. (2.15) over all possible $\mathcal{M}$, subject to $\mathcal{M}$ being a valid measurement. This results

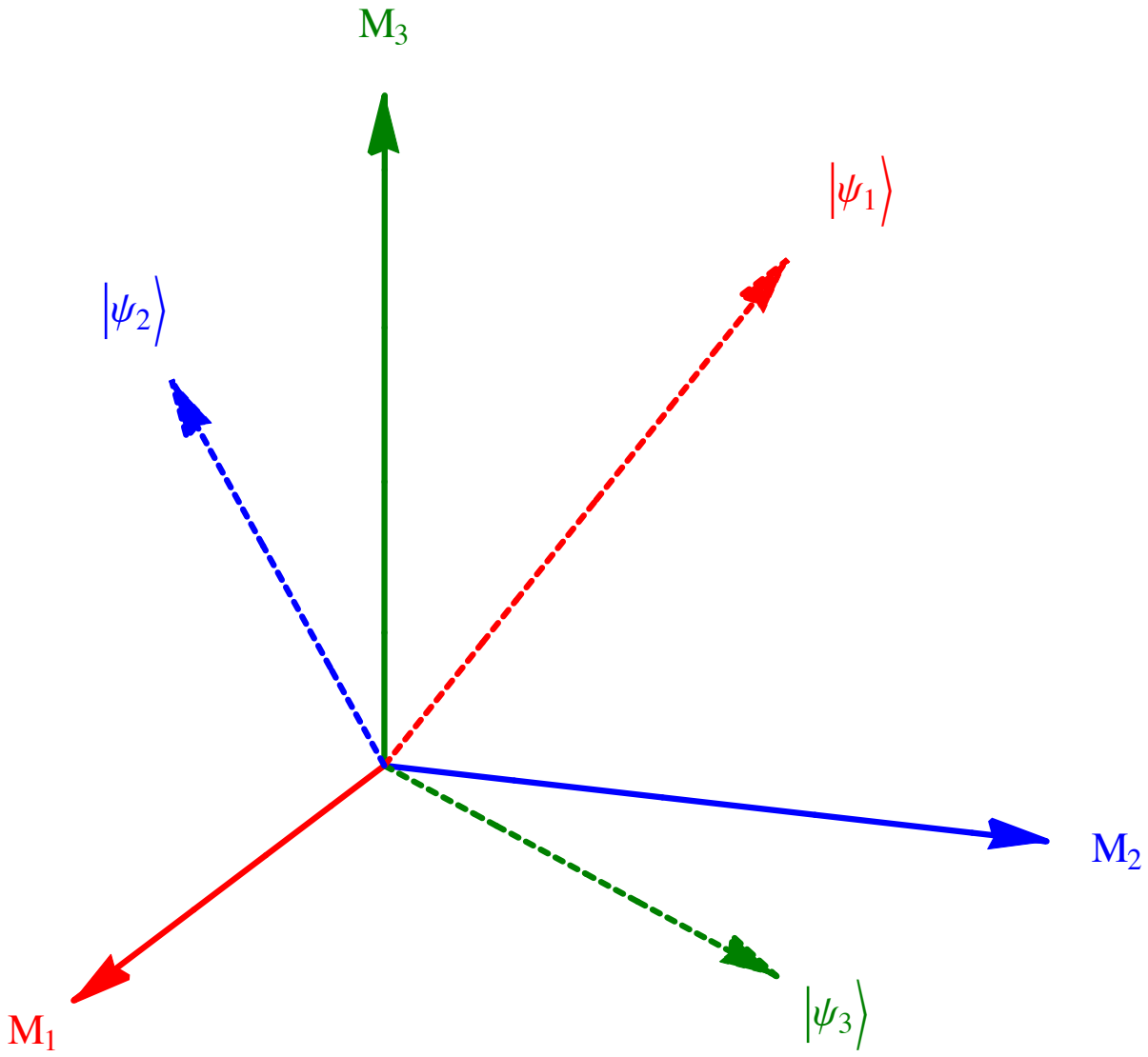



Figure 2.3: *Conclusive exclusion for pure states.* In this figure we illustrate a set of states allowing conclusive exclusion for $d = k = 3$. Here, the exclusion measurement is given by $M_1 = |1\rangle\langle 1|$, $M_2 = |2\rangle\langle 2|$ and $M_3 = |3\rangle\langle 3|$ and the states are such that $|\psi_1\rangle = a_2|2\rangle + a_3|3\rangle$, $|\psi_2\rangle = b_1|1\rangle + b_3|3\rangle$ and $|\psi_3\rangle = c_1|1\rangle + c_2|2\rangle$ with the coefficients being real and suitably normalized.

in the following (primal) SDP:

$$\begin{aligned} \underset{\mathcal{M}=\{M_i\}_{i=1}^k}{\text{Minimize:}}\ & \alpha = \sum_{i=1}^{k} \text{Tr}\left[\tilde{\rho}_i M_i\right]. \\ \text{Subject to:}\ & \sum_{i=1}^{k} M_i = \mathbb{I}, \\ & M_i \geq 0, \quad \forall i. \end{aligned} \tag{2.21}$$

As before, the constraints ensure that $\mathcal{M}$ is a valid measurement.

By comparison with the SDP for state discrimination given in Eqs. (1.20) and (1.21),[3] we know that the dual problem will be:

$$\begin{aligned} \underset{N}{\text{Maximize:}}\ & \beta = \text{Tr}\left[N\right]. \\ \text{Subject to:}\ & N \leq \tilde{\rho}_i, \quad \forall i, \\ & N \in \textit{Herm}. \end{aligned} \tag{2.22}$$

For single-state exclusion, the problem is essentially to maximize the trace of a Hermitian matrix, N, subject to $\tilde{\rho}_i - N$ being a positive semidefinite matrix for all $i$.

### 2.3.1 Optimal measurements

Applying strong duality in a similar manner as in Section 1.3.2, allows us to formulate necessary and sufficient criteria for determining whether an exclusion measurement is optimal. This leads to:

**Theorem 3.** *Suppose an unknown state, $\sigma$, is prepared from a set of known states, $\mathcal{P} = \{\rho_i\}_{i=1}^k$, according to a probability distribution, $\{p_i\}_{i=1}^k$. A measurement, $\mathcal{M} = \{M_i\}_{i=1}^k$, is optimal for attempting to exclude a preparation with minimum error if and only if:*

$$N = \sum_{i=1}^{k} \left[\tilde{\rho}_i M_i\right], \tag{2.23}$$

*is Hermitian and satisfies $N \leq \tilde{\rho}_i$, for all $i$.*

*Proof.* The proof is near identical to that of Theorem 2 so we omit it. Note that the exclusion SDP satisfies strong duality. This can be shown using Slater's theorem (Theorem 1) and taking $M_i = \frac{1}{k}\mathbb{I}$, for all $i$ and $N = -\mathbb{I}$. □

[3]Note that in contrast to the state discrimination primal problem, here we are trying to minimize rather than maximize an objective function subject to an equality constraint. This has the effect of making the dual a maximization problem and reversing the direction of the inequality constraint. Furthermore, weak duality now implies that $\alpha^{\text{opt}} \geq \beta^{\text{opt}}$.

### 2.3.2 A necessary condition for single-state conclusive exclusion

Through the application of weak duality, we can also gain insight into the feasibility of the task. As the optimal solution to the dual problem provides a lower bound on the solution of the primal problem, any feasible solution to the dual does too, although it may not necessarily be tight. This relation can be summarized as:

$$\mathrm{Tr}\left[N^{\mathrm{feas}}\right] \leq \mathrm{Tr}\left[N^{\mathrm{opt}}\right] = \beta^{\mathrm{opt}} = \alpha^{\mathrm{opt}}. \tag{2.24}$$

In particular, if, for a given $\mathcal{P}$, we can construct a feasible $N$ with $\mathrm{Tr}\left[N\right] > 0$, then we have $\alpha^{\mathrm{opt}} > 0$ and hence conclusive exclusion is not possible.

Constructing such an $N$ gives rise to the following necessary condition on the set $\mathcal{P}$ for conclusive exclusion to be possible:

**Theorem 4.** *Suppose a system is prepared in the state $\sigma$ using a preparation chosen at random from the set $\mathcal{P} = \{\rho_i\}_{i=1}^{k}$. Single-state conclusive exclusion is possible only if:*

$$\sum_{j\neq l=1}^{k} F\left(\rho_j, \rho_l\right) \leq k\left(k-2\right), \tag{2.25}$$

*where $F\left(\rho_j, \rho_l\right) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho_j}\rho_l\sqrt{\rho_j}}\right]$ is the fidelity between states $\rho_j$ and $\rho_l$.*

Before we begin the proof, note that this result captures some of our intuition about when conclusive exclusion should be possible. If the states in $\mathcal{P}$ are too similar (they have high fidelity with one another) then it is harder to tell them apart and one should not be able to conclusively exclude a state.

*Proof.* A feasible solution to the dual SDP, $N$, must be Hermitian and satisfy $N \leq \rho_i$, for all $i$. Our goal is to construct such an $N$ with the additional property that $\mathrm{Tr}\left[N\right] > 0$. If this is possible, then conclusive exclusion is not possible.

First we define $U_{jl}$ to be the unitary such that $\mathrm{Tr}\left[\sqrt{\rho_l}\sqrt{\rho_j}U_{jl}\right] = F\left(\rho_j, \rho_l\right)$,[4] and note that $U_{lj} = U_{jl}^{\dagger}$.[5] We construct $N$ as follows:

$$N = -p\sum_{r=1}^{k}\rho_r + \frac{1-\epsilon}{k-2}p\sum_{1\leq j<l\leq k}\left(\sqrt{\rho_j}U_{jl}\sqrt{\rho_l} + \sqrt{\rho_l}U_{jl}^{\dagger}\sqrt{\rho_j}\right),$$

[4] This follows from applying the polar decomposition $\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} = \sqrt{\rho}\sqrt{\sigma}U$.

[5] To see this, note that:

$$F\left(\rho_l, \rho_j\right) = F\left(\rho_j, \rho_l\right) = \mathrm{Tr}\left[\sqrt{\rho_l}\sqrt{\rho_j}U_{jl}\right] = \mathrm{Tr}\left[U_{jl}^{\dagger}\sqrt{\rho_j}\sqrt{\rho_l}\right] = \mathrm{Tr}\left[\sqrt{\rho_j}\sqrt{\rho_l}U_{jl}^{\dagger}\right].$$

where $p, \epsilon \in (0,1)$. Note that $N$ is Hermitian. Now consider:

$$\begin{aligned}
\rho_1 - N &= (1+p)\,\rho_1 + p\sum_{r=2}^{k}\rho_r - \frac{1-\epsilon}{k-2}p\sum_{1\le j<l\le k}\left(\sqrt{\rho_j}U_{jl}\sqrt{\rho_l} + \sqrt{\rho_l}U_{jl}^{\dagger}\sqrt{\rho_j}\right), \\
&= \sum_{r=2}^{k}\left[\frac{1+p}{k-1}\rho_1 + \epsilon p\rho_r - \frac{1-\epsilon}{k-2}p\left(\sqrt{\rho_1}U_{1r}\sqrt{\rho_r} + \sqrt{\rho_r}U_{1r}^{\dagger}\sqrt{\rho_1}\right)\right] \\
&\quad + \frac{1-\epsilon}{k-2}p\sum_{2\le j<l\le k}\left[\rho_j + \rho_l - \sqrt{\rho_j}U_{jl}\sqrt{\rho_l} - \sqrt{\rho_l}U_{jl}^{\dagger}\sqrt{\rho_j}\right], \\
&= \sum_{r=2}^{k}\left[\frac{1+p}{k-1}\rho_1 + \epsilon p\rho_r - \frac{1-\epsilon}{k-2}p\left(\sqrt{\rho_1}U_{1r}\sqrt{\rho_r} + \sqrt{\rho_r}U_{1r}^{\dagger}\sqrt{\rho_1}\right)\right] \\
&\quad + \frac{1-\epsilon}{k-2}p\sum_{2\le j<l\le k}\left(\sqrt{\rho_j}\sqrt{U_{jl}} - \sqrt{\rho_l}\sqrt{U_{jl}^{\dagger}}\right)\left(\sqrt{U_{jl}^{\dagger}}\sqrt{\rho_j} - \sqrt{U_{jl}}\sqrt{\rho_l}\right).
\end{aligned}$$

The terms from the second summation in the last line are positive semidefinite. Consider, individually, the terms in the first summation:

$$\begin{aligned}
&\frac{1+p}{k-1}\rho_1 + \epsilon p\rho_r - \frac{1-\epsilon}{k-2}p\left(\sqrt{\rho_1}U_{1r}\sqrt{\rho_r} + \sqrt{\rho_r}U_{1r}^{\dagger}\sqrt{\rho_1}\right), \\
&= \left[\frac{1+p}{k-1} - \left(\frac{(1-\epsilon)\,p}{k-2}\right)^2\frac{1}{\epsilon p}\right]\rho_1 \\
&\quad + \left[\left(\frac{(1-\epsilon)\,p}{k-2}\right)^2\frac{1}{\epsilon p}\right]\rho_1 + \epsilon p\rho_r - \frac{1-\epsilon}{k-2}p\left(\sqrt{\rho_1}U_{1r}\sqrt{\rho_r} + \sqrt{\rho_r}U_{1r}^{\dagger}\sqrt{\rho_1}\right), \\
&= \left[\frac{1+p}{k-1} - \left(\frac{(1-\epsilon)\,p}{k-2}\right)^2\frac{1}{\epsilon p}\right]\rho_1 \\
&\quad + \left(\frac{(1-\epsilon)\,p}{(k-2)\sqrt{\epsilon p}}\sqrt{\rho_1}\sqrt{U_{1r}} - \sqrt{\epsilon p}\sqrt{\rho_r}\sqrt{U_{1r}^{\dagger}}\right)\left(\frac{(1-\epsilon)\,p}{(k-2)\sqrt{\epsilon p}}\sqrt{U_{1r}^{\dagger}}\sqrt{\rho_1} - \sqrt{\epsilon p}\sqrt{U_{1r}}\sqrt{\rho_r}\right).
\end{aligned}$$

Hence, to guarantee that $\rho_1 - N$ is positive semidefinite, we need the first term in the last line to be positive:

$$\begin{aligned}
&\left[\frac{1+p}{k-1} - \left(\frac{(1-\epsilon)\,p}{k-2}\right)^2\frac{1}{\epsilon p}\right] \geq 0, \\
&\qquad \Rightarrow \quad \frac{\epsilon}{\frac{(k-1)(1-\epsilon)^2}{(k-2)^2} - \epsilon} \geq p. \qquad (2.26)
\end{aligned}$$

Therefore, provided $p$ and $\epsilon$ satisfy Eq. (2.26), we have $N \leq \rho_1$. Similarly, one can argue that $N \leq \rho_i$, for all $i$ and hence $N$ is a feasible solution to the dual problem.

We now wish to know under what conditions we have $\mathrm{Tr}\left[N\right] > 0$:

$$\begin{aligned} & 0 < \mathrm{Tr}\left[N\right], \\ \Rightarrow \quad & 0 < -kp + \tfrac{1-\epsilon}{k-2} p \sum_{1\leq j<l\leq k} \mathrm{Tr}\left[\sqrt{\rho_j} U_{jl} \sqrt{\rho_l} + \sqrt{\rho_l} U_{jl}^\dagger \sqrt{\rho_j}\right], \\ \Rightarrow \quad & \tfrac{k(k-2)}{1-\epsilon} < \sum_{j\neq l=1}^{k} F\left(\rho_j, \rho_l\right). \end{aligned}$$

Letting $\epsilon \to 0$ and using weak duality we obtain our result. Conclusive exclusion is not possible if $\sum_{j\neq l=1}^{k} F\left(\rho_j, \rho_l\right) > k\left(k-2\right)$. □

This is only a necessary condition for single-state conclusive exclusion, and there exist sets of states that satisfy Eq. (2.25) for which it is not possible to perform conclusive exclusion. For example, the set of states:

$$\left\{ \rho_1 = \begin{pmatrix} 1-2\epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix}, \ \rho_2 = \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & 1-2\epsilon & 0 \\ 0 & 0 & \epsilon \end{pmatrix}, \ \rho_3 = \begin{pmatrix} \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & 1-2\epsilon \end{pmatrix} \right\}, \tag{2.27}$$

satisfy Eq. (2.25) for small enough $\epsilon$ and yet are not conclusively excludable (as they all have full rank).

Nevertheless, there exist sets of states on the cusp of satisfying Eq. (2.25) for which conclusive exclusion is possible. For example, the set of states of the form:

$$|\psi_i\rangle = \sum_{j\neq i}^{k} \frac{1}{\sqrt{k-1}} |j\rangle, \tag{2.28}$$

for $i = 1$ to $k$, can be conclusively excluded by the measurement $\mathcal{M} = \{|i\rangle\langle i|\}_{i=1}^{k}$, and yet:

$$\sum_{j\neq l=1}^{k} F\left(|\psi_j\rangle, |\psi_l\rangle\right) = \sum_{j\neq l=1}^{k} |\langle\psi_j|\psi_l\rangle| = k\left(k-2\right). \tag{2.29}$$

Finally, the well known necessary condition for conclusive state discrimination can be reproduced from Theorem 4:

**Corollary 1.** *Conclusive state discrimination on the set $\mathcal{P} = \{\rho_i\}_{i=1}^{k}$ is possible only if $\mathcal{P}$ is an orthogonal set.*

As the proof from Theorem 4 is drastically longer than standard proofs of this statement (see, for example, [134]), we provide it only in Appendix A.1.

### 2.3.3 Lower bounds on the probability of error

Weak duality can also be used to obtain the following lower bounds on $\alpha^{\text{opt}}$:

**Lemma 2.** *Suppose an unknown state, $\sigma$, is prepared from a set of known states, $\mathcal{P} = \{\rho_i\}_{i=1}^k$, according to a probability distribution, $\{p_i\}_{i=1}^k$. The minimum probability of incorrectly excluding a preparation, $p_{\text{error}}$, satisfies:*

1. *Analogously to Eq.* (1.23) *in Lemma 1:*

$$p_{\text{error}} \geq d \min_i \|\tilde{\rho}_i\| . \tag{2.30}$$

2. *Given two Hermitian operators, $A$ and $B$, define:*

$$\min(A, B) = \frac{1}{2}\left[A + B - |A - B|\right]. \tag{2.31}$$

*For a permutation, $\epsilon$, acting on $k$ objects, taken from the permutation group $S_k$, consider:*

$$N_\epsilon = \min\left(\tilde{\rho}_{\epsilon(k)}, \min\left(\tilde{\rho}_{\epsilon(k-1)}, \min\left(\ldots, \min\left(\tilde{\rho}_{\epsilon(2)}, \tilde{\rho}_{\epsilon(1)}\right)\right)\right)\right). \tag{2.32}$$

*Then, analogously to Eq.* (1.29) *in Lemma 1:*

$$p_{\text{error}} \geq \max_{\epsilon \in S_k} \text{Tr}\left[N_\epsilon\right]. \tag{2.33}$$

*Proof.* The general aim is to construct an $N$ that satisfies the constraints of the dual problem given in Eq. (2.22). Then, by weak duality, we have that $p_{\text{error}} \geq \text{Tr}\left[N\right]$.

1. Let $\lambda^{\min} = \min_i \|\tilde{\rho}_i\|$. To obtain the bound, define:

$$N = \lambda^{\min} \mathbb{I}_d.$$

Note that taking the trace of $N$ gives Eq. (2.30). To see that $N \leq \tilde{\rho}_i$, consider:

$$\begin{aligned} \tilde{\rho}_i - N &= \sum_{j=1}^{d} \left(\lambda_j^i - \lambda^{\min}\right) |u_j^i\rangle\langle u_j^i|, \\ &\geq 0, \end{aligned}$$

where $\sum_{j=1}^d \lambda_j^i |u_j^i\rangle\langle u_j^i|$ is the spectral decomposition of $\tilde{\rho}_i$ and we have written $\mathbb{I}_d = \sum_{j=1}^d |u_j^i\rangle\langle u_j^i|$.

2. Note that $\min(A, B) \leq A$ and $\min(A, B) \leq B$ as:

$$\begin{aligned} A - \min(A, B) &= \frac{1}{2}\left[A - B + |A - B|\right], \\ &= \frac{1}{2}\left[\sum_{i=1}^{d} \lambda_i |u_i\rangle\langle u_i| + \sum_{i=1}^{d} |\lambda_i|\, |u_i\rangle\langle u_i|\right], \\ &\geq 0, \end{aligned}$$

where $\sum_{i=1}^{d} \lambda_i |u_i\rangle\langle u_i|$ is the spectral decomposition of $A - B$.

To obtain the bound, construct $N_\epsilon$ iteratively as follows:

$$\begin{aligned} N_2 &= \min\left(\tilde{\rho}_{\epsilon(2)}, \tilde{\rho}_{\epsilon(1)}\right) \\ N_3 &= \min\left(\tilde{\rho}_{\epsilon(3)}, N_2\right) \\ &\vdots = \vdots \\ N_\epsilon = N_k &= \min\left(\tilde{\rho}_{\epsilon(k)}, N_{k-1}\right). \end{aligned}$$

Using the fact that $\min(A, B) \leq A$ and $\min(A, B) \leq B$, by construction we have $N_\epsilon \leq \tilde{\rho}_i$. By maximizing over all $\epsilon \in S_k$, we obtain Eq. (2.33).

□

## 2.4 Aside: Alternative formulations of state exclusion

As noted in Chapter 1, there exist multiple strategies and figures of merit when undertaking state discrimination. Similarly, alternative targets to minimum error can be defined in the problem of state exclusion. In this section, we define two of these, unambiguous and worst-case error exclusion, and construct the related SDPs.

### 2.4.1 Unambiguous state exclusion

In *unambiguous state exclusion* on the set of preparations $\mathcal{P} = \{\tilde{\rho}_i\}_{i=1}^{k}$, we consider a measurement given by $\mathcal{M} = \{M_1, \ldots, M_k, M_?\}$. If we obtain measurement outcome $i \in \{1, \ldots, k\}$, then we can exclude with certainty the state $\rho_i$. However, if we obtain the outcome labeled '?', we cannot infer which state to exclude. We wish to minimize the probability of obtaining this inconclusive measurement:

$$\alpha = \sum_{i=1}^{k} \mathrm{Tr}\left[\tilde{\rho}_i M_?\right], \tag{2.34}$$

which can be rewritten as:

$$\alpha = \mathrm{Tr}\left[\sum_{j=1}^{k} \tilde{\rho}_j \left(\mathbb{I} - \sum_{i=1}^{k} M_i\right)\right]. \tag{2.35}$$

Defining $\tilde{\alpha} = 1 - \alpha$, the primal SDP associated with this task is given by:

$$\begin{aligned} \underset{\{M_i\}_{i=1}^k}{\text{Maximize:}}\ & \tilde{\alpha} = \mathrm{Tr}\left[\sum_{j=1}^{k} \tilde{\rho}_j \sum_{i=1}^{k} M_i\right]. \\ \text{Subject to:}\ & \sum_{i=1}^{k} M_i \leq \mathbb{I}, \\ & \mathrm{Tr}\left[\tilde{\rho}_i M_i\right] = 0, \quad 1 \leq i \leq k, \\ & M_i \geq 0, \quad 1 \leq i \leq k. \end{aligned} \tag{2.36}$$

Here, the first and third constraints ensure that $\mathcal{M}$ is a valid measurement whilst the second, $\mathrm{Tr}\left[\tilde{\rho}_i M_i\right] = 0$, $1 \leq i \leq k$, encapsulates the fact that when measurement outcome $i$ occurs we should be able to exclude state $\rho_i$ with certainty.

The dual problem can be shown to be (see Appendix B.1):

$$\begin{aligned} \underset{N,\{a_i\}_{i=1}^k}{\text{Minimize:}}\ & \beta = \mathrm{Tr}\left[N\right]. \\ \text{Subject to:}\ & a_i \tilde{\rho}_i + N \geq \sum_{j=1}^{k} \tilde{\rho}_j, \quad 1 \leq i \leq k, \\ & a_i \in \mathbb{R}, \quad \forall i, \\ & N \geq 0. \end{aligned} \tag{2.37}$$

Unambiguous state exclusion has recently found use in implementations of quantum digital signatures [50], enabling such schemes to be put into practice without the need for long term quantum memory.

### 2.4.2 Worst-case error state exclusion

The goal of the SDP given in Eqs. (2.21) and (2.22) is to minimize the average probability of error, over all possible preparations, of the strategy, 'if outcome $j$ occurs say $\sigma \neq \rho_j$'. An alternative goal would be to minimize the worst-case probability of error that occurs:

$$\alpha = \max_i \mathrm{Tr}\left[\tilde{\rho}_i M_i\right]. \tag{2.38}$$

The primal SDP associated with this task is:

$$\begin{aligned}
&\underset{\mathcal{M}=\{M_i\}_{i=1}^k}{\text{Minimize:}} \ \alpha = \lambda. \\
&\text{Subject to:} \ \lambda \geq \text{Tr}\left[\tilde{\rho}_i M_i\right], \quad \forall i, \\
&\qquad \sum_{i=1}^{k} M_i = \mathbb{I}, \\
&\qquad \lambda \geq 0 \in \mathbb{R}, \\
&\qquad M_i \geq 0, \quad 1 \leq i \leq k.
\end{aligned} \tag{2.39}$$

These constraints again encode that $\mathcal{M}$ forms a valid measurement and ensure that $\alpha$ picks out the worst-case error probability across all possible preparations.

The associated dual problem is given by:

$$\begin{aligned}
&\underset{N,\{a_i\}_{i=1}^k}{\text{Maximize:}} \ \beta = \text{Tr}\left[N\right]. \\
&\text{Subject to:} \ N \leq a_i \tilde{\rho}_i, \quad \forall i, \\
&\qquad \sum_{i=1}^{k} a_i \leq 1, \\
&\qquad a_i \geq 0 \in \mathbb{R}, \quad \forall i, \\
&\qquad N \in \textit{Herm}.
\end{aligned} \tag{2.40}$$

The derivation of this is given in Appendix B.2.

## 2.5 Applications of the state exclusion SDP

### 2.5.1 Optimality of the PBR measurement

As a first application of the state exclusion SDP, we now prove that the condition that $2^{1/n} \leq \tan\left(\frac{\theta}{2}\right)$ given in Eq. (2.10), is necessary as well as sufficient for being able to perform conclusive exclusion on the set of PBR states:

$$\mathcal{P}\left(\theta\right) = \left\{ \left|\Psi_{\vec{x}}\left(\theta\right)\right\rangle = \bigotimes_{i=1}^{n} \left|\psi_{x_i}\left(\theta\right)\right\rangle \right\}_{\vec{x}\in\{0,1\}^n}, \tag{2.41}$$

where $\left|\psi_0\left(\theta\right)\right\rangle$ and $\left|\psi_1\left(\theta\right)\right\rangle$ are defined as per Eq. (2.7) and $0 \leq \theta \leq \pi/2$. These states, and knowledge of the smallest value of $\theta$ for which conclusive exclusion is possible for given $n$, will be vital for the communication tasks we consider in Chapter 3.

Note that $|\Psi_{\vec{x}}(\theta)\rangle$ can be expanded to give:

$$|\Psi_{\vec{x}}(\theta)\rangle = \sum_{\vec{r}\in\{0,1\}^n} (-1)^{\vec{x}\cdot\vec{r}} \left[\cos\left(\frac{\theta}{2}\right)\right]^{n-|\vec{r}|} \left[\sin\left(\frac{\theta}{2}\right)\right]^{|\vec{r}|} |\vec{r}\rangle, \tag{2.42}$$

where $|\vec{r}| = \sum_{i=1}^{n} r_i$. Numerical solutions to the SDP (performed using [114, 160]) suggest that when $n$ and $\theta$ are such that $2^{1/n} \geq \tan\left(\frac{\theta}{2}\right)$, the measurement to perform is given by $\mathcal{M} = \{|\zeta_{\vec{x}}\rangle\langle\zeta_{\vec{x}}|\}_{\vec{x}\in\{0,1\}^n}$, where:

$$|\zeta_{\vec{x}}\rangle = \frac{1}{\sqrt{2^n}} \left(|\vec{0}\rangle - \sum_{\vec{r}\neq\vec{0}} (-1)^{\vec{x}\cdot\vec{r}} |\vec{r}\rangle\right). \tag{2.43}$$

Note that the measurement is independent of $\theta$. This insight leads to the following result:

**Theorem 5.** *Consider the set of PBR states $\mathcal{P}(\theta)$ as defined in Eq. (2.41), each prepared with equal probability. Then provided $0 \leq \theta \leq 2\arctan\left(2^{1/n} - 1\right)$:*

1. *$\mathcal{M} = \{|\zeta_{\vec{x}}\rangle\langle\zeta_{\vec{x}}|\}_{\vec{x}\in\{0,1\}^n}$, with $|\zeta_{\vec{x}}\rangle$ defined as per Eq. (2.43), is the optimal measurement for attempting to perform conclusive state exclusion.*

2. *The minimum probability of error achievable in attempting to perform conclusive state exclusion is given by:*

$$p_{error} = \frac{1}{2^n} \left[\cos\left(\frac{\theta}{2}\right)\right] \left(2 - \left[1 + \tan\left(\frac{\theta}{2}\right)\right]^n\right)^2. \tag{2.44}$$

*Hence, conclusive state exclusion is not possible for $0 \leq \theta < 2\arctan\left(2^{1/n} - 1\right)$.*

*Proof.* The proof shall be split into three parts. Firstly, we shall show that $\mathcal{M}$ is in fact a measurement before proving that it is optimal for the exclusion task considered here. Finally, we derive how well it performs at the task to obtain Eq. (2.44).

We begin by showing that $\mathcal{M}$ is a valid measurement as the projectors are all orthonormal. Consider:

$$\begin{aligned}
\langle\zeta_{\vec{s}}|\zeta_{\vec{t}}\rangle &= \frac{1}{2^n} \left(\langle\vec{0}| - \sum_{\vec{r}\neq\vec{0}} (-1)^{\vec{s}\cdot\vec{r}} \langle\vec{r}|\right) \left(|\vec{0}\rangle - \sum_{\vec{q}\neq\vec{0}} (-1)^{\vec{t}\cdot\vec{q}} |\vec{q}\rangle\right) \\
&= \frac{1}{2^n} \left(1 + \sum_{\vec{r},\vec{q}\neq\vec{0}} (-1)^{\vec{s}\cdot\vec{r}} (-1)^{\vec{t}\cdot\vec{q}} \langle\vec{r}|\vec{q}\rangle\right) \\
&= \frac{1}{2^n} \sum_{\vec{r}\in\{0,1\}^n} (-1)^{(\vec{s}+\vec{t})\cdot\vec{r}} \\
&= \delta_{\vec{s}\vec{t}}.
\end{aligned}$$

Hence $\mathcal{M}$ is a set of orthogonal vectors and therefore a valid measurement basis.

To show that this measurement is optimal for certain pairs of $n$ and $\theta$, we make use of Theorem 3, constructing an $N$ as per Eq. (2.23) and showing that it satisfies the constraints of the dual problem. Omitting the label $\theta$ from the states for brevity and writing $\tilde{\rho}_{\vec{x}} = \frac{1}{2^n}|\Psi_{\vec{x}}\rangle\langle\Psi_{\vec{x}}|$ and $M_{\vec{x}} = |\zeta_{\vec{x}}\rangle\langle\zeta_{\vec{x}}|$, we have:

$$N = \frac{1}{2^n}\sum_{\vec{x}} |\Psi_{\vec{x}}\rangle\langle\Psi_{\vec{x}}|\zeta_{\vec{x}}\rangle\langle\zeta_{\vec{x}}|.$$

Note that:

$$\begin{aligned}\langle\Psi_{\vec{x}}|\zeta_{\vec{x}}\rangle &= \frac{1}{\sqrt{2^n}}\left(\left[\cos\left(\frac{\theta}{2}\right)\right]^n - \sum_{i=1}^{n}\binom{n}{i}\left[\cos\left(\frac{\theta}{2}\right)\right]^{n-i}\left[\sin\left(\frac{\theta}{2}\right)\right]^i\right), \\ &= \frac{1}{\sqrt{2^n}}\left[\cos\left(\frac{\theta}{2}\right)\right]^n\left(2 - \left[1+\tan\left(\frac{\theta}{2}\right)\right]^n\right).\end{aligned}$$

So we have:

$$N = C(\theta)\left[|\vec{0}\rangle\langle\vec{0}| - \sum_{\vec{r}\neq\vec{0}}\left[\tan\left(\frac{\theta}{2}\right)\right]^{|\vec{r}|}|\vec{r}\rangle\langle\vec{r}|\right],$$

where $C(\theta)$ is given by:

$$C(\theta) = \frac{1}{2^n}\left[\cos\left(\frac{\theta}{2}\right)\right]^{2n}\left(2 - \left[1+\tan\left(\frac{\theta}{2}\right)\right]^n\right).$$

Note also that $N$ is a real, diagonal matrix and hence is Hermitian so it remains to determine under what conditions $\tilde{\rho}_{\vec{x}} - N$ is a positive semidefinite matrix for all $\vec{x}$. If $\theta = 2\arctan\left(2^{1/n}-1\right)$, then $N = 0$ and this is satisfied trivially so, in what follows, we restrict our attention to $0 \leq \theta < 2\arctan\left(2^{1/n}-1\right)$.

Let us define the matrices $A_{\vec{x}}$ by:

$$A_{\vec{x}} = -N + \tilde{\rho}_{\vec{x}}.$$

The goal is to prove that none of the $A_{\vec{x}}$ have a negative eigenvalue. Say $A_{\vec{x}}$ has eigenvalues $\left\{a^r_{\vec{x}}\right\}_{r=1}^{2^n}$, where $a^1_{\vec{x}} \geq a^2_{\vec{x}} \geq \ldots \geq a^{2^n}_{\vec{x}}$. The matrix $-N$ has eigenvalues $\{v^r\}_{r=1}^{2^n}$, where for $1 \leq r \leq 2^n - 1$:

$$v^r = C(\theta)\left[\tan\left(\frac{\theta}{2}\right)\right]^{|\vec{r}|},$$

and for $r = 2^n$:

$$v^{2^n} = -C(\theta).$$

Each $\tilde{\rho}_{\vec{x}}$ is a rank 1 subnormalized density matrix and hence they have eigenvalues $u_{\vec{x}}^1 = \frac{1}{2^n}$ and $u_{\vec{x}}^r = 0$ for $2 \leq r \leq 2^n$.

By Weyl's inequality [168, 87]:

$$v^r + u_{\vec{x}}^{2^n} \leq a_{\vec{x}}^r.$$

So, provided $C(\theta) > 0$, we have $a_{\vec{x}}^r > 0$ for $1 \leq r \leq 2^n - 1$. Hence at most one eigenvalue of $A_{\vec{x}}$ is non-positive. Investigating this non-positive eigenvalue further, consider $A_{\vec{x}}$ acting on the state $|\zeta_{\vec{x}}\rangle$:

$$\begin{aligned} A_{\vec{x}}|\zeta_{\vec{x}}\rangle &= \tilde{\rho}_{\vec{x}}|\zeta_{\vec{x}}\rangle - \sum_{\vec{y}\in\{0,1\}^n} \tilde{\rho}_{\vec{y}}|\zeta_{\vec{y}}\rangle\langle\zeta_{\vec{y}}|\zeta_{\vec{x}}\rangle, \\ &= 0. \end{aligned}$$

Hence the non-positive eigenvalue of $A_{\vec{x}}$ is 0, implying that $A_{\vec{x}} \geq 0$, for all $\vec{x}$, which in turn implies that $N \leq \rho_{\vec{x}}$, for all $\vec{x}$, provided $C(\theta) > 0$. As $[\cos(\theta/2)]^{2n} \geq 0$, we have shown that $\mathcal{M} = \{|\zeta_{\vec{x}}\rangle\langle\zeta_{\vec{x}}|\}_{\vec{x}\in\{0,1\}^n}$ is the optimal measurement for exclusion provided:

$$\left(2 - \left[1 + \tan\left(\frac{\theta}{2}\right)\right]^n\right) \geq 0.$$

Lastly, to obtain Eq. (2.44) we take the trace of $N$. This leads to:

$$\mathrm{Tr}[N] = \frac{1}{2^n}\left[\cos\left(\frac{\theta}{2}\right)\right]^{2n}\left(2 - \left[1 + \tan\left(\frac{\theta}{2}\right)\right]^n\right)^2,$$

which is strictly positive provided $\left(2 - \left[1 + \tan\left(\frac{\theta}{2}\right)\right]^n\right) > 0$ and hence, in this region, conclusive exclusion is not possible. □

### 2.5.2 Measures of state assignment compatibility

A second application of the state exclusion SDP is with regards to *state compatibility*. Suppose $k$ parties, labeled by $i$, each consider the state of the same quantum system and assign to it a description given by the density matrix $\rho_i$. Such a scenario could occur, for example, because:

1. Each party was given differing or incomplete information on the state the system was prepared in.

2. Each party may hold part of a multipartite entangled state which is also entangled with an ancilla system that no party has access to [28]. If each party measures their part of the state, and does not reveal their measurement choice or outcome to the other parties, then in general they will assign different reduced density matrices to the ancilla.

A natural question to ask in either scenario is under what conditions are the states the parties assign to the system compatible? Is it possible for them to deduce that one of their assignments is definitely wrong as their descriptions are contradictory? Furthermore, if they are regarded as compatible, then one might wonder if it is possible to assign a degree of confidence to this statement.

This last consideration leads to the notion of a measure for how compatible a set of state descriptions are. For example, for $k = 3$ and with respect to most criteria, the set of states given in Eq. (2.27) will be regarded as compatible assignments provided $\epsilon > 0$ even though for small $\epsilon$ they are practically orthogonal and nearly contradictory. A measure of compatibility seeks to capture this.

There are numerous definitions of what it actually means for a set of assignments to be compatible [39] but here we will focus on three in particular and show how each can be quantified using an SDP.

### Post-Peierls compatibility

We begin with a compatibility criteria closely related to state exclusion - *post-Peierls (PP) compatibility* [136, 39]. It is based on trying to perform a measurement on the system such that, from the outcome, it is always possible to deduce that one of the parties' state assignments is incorrect. This happens if the outcome produced was such that a party had assigned zero probability to its occurrence. If such a measurement exists, the assignments are PP incompatible and an assignment can be ruled out in a single-shot process.

**Definition 1** (PP compatibility)**.** *Given a set of $k$ assignments, $\mathcal{P} = \{\rho_i\}_{i=1}^{k}$, for the state of a $d$-dimensional quantum system, we say that $\mathcal{P}$ is PP compatible if and only if for all measurements, $\mathcal{M} = \{M_j\}$, there exists an outcome, $j$, such that:*

$$\mathrm{Tr}\left[M_j \rho_i\right] > 0, \quad \forall i. \tag{2.45}$$

Note that conclusive exclusion is possible for $\mathcal{P}$, if and only if $\mathcal{P}$ is PP incompatible. Given this relation to state exclusion, a natural measure of the PP compatibility of a set is given by the probability of error in attempting to exclude them. This measure can be formulated as a slight variation on the state exclusion SDP in Eq. (2.22):

**Definition 2** (Measure of PP compatibility)**.** *The measure of the PP compatibility of a set $\mathcal{P}$*

*is denoted by* $\mathcal{K}_{PP}(\mathcal{P})$ *and defined as the solution to:*

$$\begin{aligned} &\underset{N}{\textit{Maximize:}} && \mathrm{Tr}\left[N\right]. \\ &\textit{Subject to:} && N \leq \rho_i, \quad \forall i, \\ & && N \in \textit{Herm}. \end{aligned} \tag{2.46}$$

The associated dual SDP is, in analogy to Eq. (2.21):

$$\begin{aligned} &\underset{\mathcal{M}=\{M_i\}_{i=1}^k}{\text{Minimize:}} && \sum_{i=1}^{k} \mathrm{Tr}\left[\rho_i M_i\right]. \\ &\text{Subject to:} && \sum_{i=1}^{k} M_i = \mathbb{I}, \\ & && M_i \geq 0, \quad \forall i, \end{aligned} \tag{2.47}$$

and by strong duality, the result of both optimizations given in Eqs. (2.46) and (2.47) will be the same. Note that $\mathcal{K}_{\mathrm{PP}}(\mathcal{P})$ has the desirable properties that $\mathcal{K}_{\mathrm{PP}}(\mathcal{P}) = 0$ iff $\mathcal{P}$ is PP incompatible and $\mathcal{K}_{\mathrm{PP}}(\mathcal{P}) = 1$ iff all state assignments in $\mathcal{P}$ are identical.

### Brun-Finkelstein-Mermin compatibility

*Brun-Finkelstein-Mermin (BFM) compatibility* [28] deals with the prior beliefs of the $k$ parties rather than the possibility of contradictory measurement outcomes and was originally formulated to deal with the entangled state scenario detailed above. If parties hold BFM compatible state assignments, then there exists a density matrix that does not contradict any of the parties' beliefs: no measurement performed on such a density matrix will produce an outcome a party assigns zero probability to.

**Definition 3** (BFM compatibility)**.** *Given a set of* $k$ *assignments,* $\mathcal{P} = \{\rho_i\}_{i=1}^k$*, for the state of a* $d$*-dimensional quantum system, we say that* $\mathcal{P}$ *is BFM compatible if and only if:*

$$\bigcap_{i=1}^{k} \mathrm{supp}\left(\rho_i\right) \neq \emptyset. \tag{2.48}$$

If states are BFM compatible, then they are also PP compatible. To see this consider a measurement, $\mathcal{M} = \{M_j\}$, and note that if the states are BFM compatible, the intersection of their supports is nonempty. As the $\{M_j\}$ span the Hilbert space, there exists an $M_j$ such that $\bigcap_{i=1}^{k} \mathrm{supp}\left(\rho_i\right) \cap \mathrm{supp}\left(M_j\right) \neq \emptyset$ and hence $\mathrm{Tr}\left[M_j \rho_i\right] > 0$, for all $i$. This implies that conclusive exclusion is not possible for $\mathcal{P}$ and that the set of states must be PP compatible.

To quantify the degree of BFM compatibility of a set of states, we adapt a measure suggested by Kitaev [103]. The idea is to find a positive matrix, $R$, that 'fits' into the density matrices in $\mathcal{P}$ to the greatest possible degree. Rephrasing this as an SDP we obtain:

**Definition 4** (Measure of BFM compatibility)**.** *The measure of the BFM compatibility of a set $\mathcal{P}$ is denoted by $\mathcal{K}_{BFM}(\mathcal{P})$ and defined as the solution to:*

$$\begin{aligned} \underset{R}{\textit{Maximize:}}\ & \mathrm{Tr}\left[R\right]. \\ \textit{Subject to:}\ & R \leq \rho_i, \quad \forall i, \\ & R \geq 0. \end{aligned} \tag{2.49}$$

This is similar in form to Eq. (2.46) and hence it is easy to see that the related dual problem is:

$$\begin{aligned} \underset{\mathcal{M}=\{M_i\}_{i=1}^k}{\text{Minimize:}}\ & \sum_{i=1}^{k} \mathrm{Tr}\left[\rho_i M_i\right]. \\ \text{Subject to:}\ & \sum_{i=1}^{k} M_i \geq \mathbb{I}, \\ & M_i \geq 0, \quad \forall i. \end{aligned} \tag{2.50}$$

As $R$ is constrained to be positive semidefinite, this transforms the equality constraint in the dual into an inequality. Once again, the SDP satisfies strong duality and $\mathcal{K}_{\mathrm{BFM}}(\mathcal{P}) = 0$ for BFM incompatible states whilst $\mathcal{K}_{\mathrm{BFM}}(\mathcal{P}) = 1$ if all of the state assignments are identical.

### Equal support compatibility

A more restrictive condition than BFM, termed *equal support (ES) compatibility* [39], is, like the PP criteria, based on the compatibility of measurement outcomes. Here, states are incompatible if there exists a measurement that, if one were given access to an unlimited number of copies of the system, applying it to each copy individually would eventually lead to an outcome that contradicts one of the parties' assignments.

**Definition 5** (ES compatibility)**.** *Given a set of $k$ assignments, $\mathcal{P} = \{\rho_i\}_{i=1}^k$, for the state of a $d$-dimensional quantum system, we say that $\mathcal{P}$ is ES compatible if and only if for all measurements, $\mathcal{M} = \{M_j\}$, and for each measurement outcome, $j$, either:*

$$\begin{aligned} & \mathrm{Tr}\left[M_j \rho_i\right] > 0, \quad \forall i, \\ \textit{or}\quad & \mathrm{Tr}\left[M_j \rho_i\right] = 0, \quad \forall i. \end{aligned} \tag{2.51}$$

If parties hold ES compatible state assignments, then each state has the same support. Hence, it is easy to see that if a set of states is ES compatible, then it is also BFM compatible.

Again, we can measure the degree of compatibility using an SDP:

**Definition 6** (Measure of ES compatibility)**.** *The measure of the ES compatibility of a set $\mathcal{P}$ is denoted by $\mathcal{K}_{ES}(\mathcal{P})$ and defined as the solution to:*

$$\begin{aligned} &\underset{\lambda}{\textit{Maximize:}}\ \lambda. \\ &\textit{Subject to:}\ \sum_{j=1}^{k} \lambda \rho_j \leq \rho_i, \quad \forall i, \\ &\qquad\qquad \lambda \geq 0. \end{aligned} \tag{2.52}$$

It can be shown (see Appendix B.3) that the dual SDP is:

$$\begin{aligned} \underset{\{\alpha_i\}_{i=1}^{d}, \mathcal{M}=\{M_i\}_{i=1}^{k}}{\text{Minimize:}} \ & \sum_{i=1}^{k} \text{Tr}\left[\rho_i M_i\right], \\ \text{Subject to:}\ & \sum_{i=1}^{d} \alpha_i \geq 1, \\ & \sum_{j=1}^{k} \rho_j \sum_{i=1}^{k} M_i \geq \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_d \end{pmatrix}, \\ & \alpha_i \in \mathbb{R}_0^+, \\ & M_i \geq 0, \ \forall i. \end{aligned} \tag{2.53}$$

As before, strong duality is satisfied and the measure displays the desired properties for both incompatible and identical sets of states.

### The relationship between the compatibility measures

As shown in [39] and mentioned here, the compatibility criteria form a hierarchy. For a set of state assignments:

$$\text{ES compatible} \Rightarrow \text{BFM compatible} \Rightarrow \text{PP compatible}. \tag{2.54}$$

The compatibility measures defined above also reflect this ordering. This is captured in the following theorem:

**Theorem 6.** *Given a set of k assignments, $\mathcal{P} = \{\rho_i\}_{i=1}^k$, for the state of a d-dimensional quantum system, the following holds:*

$$\mathcal{K}_{PP}(\mathcal{P}) \geq \mathcal{K}_{BFM}(\mathcal{P}) \geq \mathcal{K}_{ES}(\mathcal{P}). \tag{2.55}$$

*Proof.* The goal is to show that the optimal solution to one SDP can be manipulated to give a feasible to solution to another.

First, consider the relation between PP and BFM. An $R$ satisfying the constraints of Eq. (2.49) will also satisfy the constraints of Eq. (2.46) as being positive semidefinite implies that $R$ is Hermitian. Hence, if $R^{\text{opt}}$ is the optimal solution to Eq. (2.49) and $N^{\text{opt}}$ is the optimal solution to Eq. (2.46), then $\text{Tr}\left[N^{\text{opt}}\right] \geq \text{Tr}\left[R^{\text{opt}}\right]$ and the first inequality in Eq. (2.55) holds.

Finally, consider the relation between BFM and ES. Suppose $\mathcal{M}$ is a feasible solution to the BFM dual SDP given in Eq. (2.50). Then, by picking the set $\{\alpha_i\}_{i=1}^d$ to be the eigenvalues of $\sum_{j=1}^k \rho_j$, we obtain $\left(\{\alpha_i\}_{i=1}^d, \mathcal{M}\right)$ as a solution to the dual SDP for ES in Eq. (2.53). As the objective function for both SDPs is $\sum_{i=1}^k \text{Tr}\left[\rho_i M_i\right]$, the optimum value for the BFM dual SDP can also be obtained in the ES dual SDP. Remembering that in these dual problems we are minimizing with respect to the constraints, this implies the second inequality in Eq. (2.55). □

## 2.6 Summary

In this chapter we explored the possibility of $\psi$-epistemic models explaining the indistinguishability of quantum states within the framework of ontological models. Such models seemingly have the potential to explain many of quantum mechanics interesting properties as illustrated by Spekkens' toy model. However, the no-go theorem of Pusey, Barrett and Rudolph shows that, subject to assumptions, a $\psi$-epistemic ontological model is not able to fully reproduce the predictions of quantum theory.

The proof of the PBR result motivated the investigation of a close cousin to the problem of state discrimination. We termed this state exclusion as it concerns excluding a preparation from a given set that may have taken place. Like its relative, state exclusion can be formulated as an SDP which proved useful for determining conditions for measurements to be optimal, bounds on the probability of success and a necessary condition for it to be achievable conclusively.

Using the state exclusion SDP we were able to define measures for various criteria regarding state compatibility and investigate the relationship between them. The SDP, together with results derived from it, also proved useful in determining the optimality of the PBR proof.

There are many more areas still to explore with respect to state exclusion. One avenue of potential research would be to determine bounds on the probability of error in addition to those given in Lemma 2. In particular, it would be useful to have upper bounds on this probability. One approach would be to use the equivalency with state discrimination discussed in Section 2.2.1, to transfer known bounds between the two scenarios. Intuitively it would also seem that for a set of $k$ pure states for which $|\langle\psi_i|\psi_j\rangle| \leq \beta$, for all $i \neq j$, there should exist a constant, $\tilde{\beta}$, such that if $\beta \leq \tilde{\beta}$, then conclusive exclusion is possible. However, this intuition is subject to a note of caution: some measures for the distinguishability of a set of states do not increase as the pairwise overlap between elements in the set decreases [101].

It is an open question, originally posed in [39], as to whether a projective measurement is always optimal for attempting to perform conclusive exclusion on a set of linearly independent pure states. For state discrimination, the optimal measurement is known to be such that the rank of the measurement operator is less than the rank of the corresponding density matrix [63]. Attempting a similar proof here only yields that a projective measurement is certainly optimal when conclusive exclusion is not possible to the extent that $\mathrm{Tr}\left[M_i|\psi_i\rangle\langle\psi_i|\right] > 0$, for all $i$. A proof of this is found in Appendix A.2.

Finally, analogues to Theorems 3 and 4 and Lemma 2 for unambiguous and worst-case error state exclusion are still to be derived. With respect to the former of these, such results may find application in analyzing implementations of digital signatures.

The PBR theorem is arguably the most significant result in the foundations of quantum theory in recent years. Other foundational results, such as Bell's theorem, have provided routes to the discovery of the power of quantum mechanics with respect to its classical counterpart. Can similar consequences be unearthed using the PBR result? As we shall see in the next chapter, the answer is a dramatic yes.

# Chapter 3

# Communication Tasks with Infinite Quantum-Classical Separation

## 3.1 Communication tasks and protocols

In a typical communication task, two players, Alice and Bob, are given inputs $x$ and $y$ and asked to compute some function or relation, $f(x, y)$. As initially neither player has any knowledge of the other's input, some communication will have to take place between the parties to achieve their goal. Depending on the resources available to them, this communication may involve sending quantum states or perhaps be restricted to sending classical messages. How much of an advantage can be gained in using quantum resources over classical ones? The standard measure used to investigate this question is the communication complexity [173], the minimum number of bits or qubits the players must exchange to succeed. Tasks exist for which there is an exponential separation between the quantum and classical communication complexities including Raz's eponymous problem [145], quantum fingerprinting [32] and the vector in the subspace problem [147]. Indeed, in the absence of shared entanglement, it is known that such a separation is maximal in the bounded error model [107].

Here we shall consider two alternative resources and ask how big the separation can be. Firstly, rather than analyzing how much communication is needed in a given task, one can look at how much information the players need to exchange regarding their inputs. More formally this is referred to as the *internal information cost* of the protocol.[1] The information cost is a

[1] One can also define a quantity called the *external information cost*: the amount of information the players reveal to an external observer. However, in most of what we consider here, the players' inputs shall come from

useful quantity as it lower bounds a protocol's communication complexity [40, 11]. In classical information theory, it has found use in proving direct sum theorems [40, 11, 94, 12] and while for quantum protocols involving multiple rounds there have been many definitions (see, for example, [95, 26, 93] and in particular [162] for a recent, fully quantum generalization of the classical case), it is relatively simple to define for single round schemes and these will be sufficient for our current purposes. If one wants to reveal as little information as possible, it is natural to ask if an advantage can be gained in using quantum protocols instead of classical ones and there are known exponential separations [102]. One can also consider the separations that can be achieved between the information cost and the communication complexity of a task. In classical information theory, for constant non-zero error, the gap is at most exponential [26, 72]. For zero-error, the largest known gap is constant against linear and occurs for the equality function [26].

Our second alternative will be to consider the communication complexity when the players are allowed to share an unlimited amount of entanglement in the quantum setting. This scenario was originally formulated in [45] (and developed in [35, 31]) where a task was found for which sharing an entangled state reduces the communication complexity by a single bit. Exponential separations between what is possible with entanglement assisted and classical strategies have also been found [73, 75] but in general, it is known that almost all Boolean functions have linear communication complexity even in the presence of shared entanglement [34, 74, 126]. A recent survey of the field can be found in [30].

Here, we use the PBR result to design a communication task that results in beyond-exponential separations between the relative power of using quantum and classical resources. With respect to the information cost, we find that in the zero-error setting it is possible to have an infinite separation:[2] classically nearly all of the information needs to be revealed while a quantum strategy can succeed and yet reveal next to nothing. This result has clear implications if one is concerned about keeping such information private. If we instead want a separation with respect to the number of sent bits, rather than the amount of sent information, we are able to do so by allowing the players to abort some fraction of the games they play. Here an entanglement assisted strategy has constant communication complexity, while any purely clas-

a product distribution and under these circumstances, the internal and external information costs are equal.

[2]We shall refer to a separation between the classical and quantum versions of a complexity measure or cost as *infinite* if the classical version tends to infinity as the problem size increases and the corresponding quantum measure tends to a constant. For a more complete description, see Definitions 8 and 9.

sical strategy has complexity linear in the problem size. In the absence of shared entanglement, we shall also be able to upper and lower bound the number of qubits that must be sent in the task and this will lead to a separation between the quantum information and communication costs, qualitatively different to those known for their classical counterparts.

Such unbounded separations for communication tasks have previously been found in the nondeterministic setting. Here, for a Boolean function $f$, two parties are required to compute $f$ correctly with certainty if $f(x,y) = 0$ and with non-zero probability if $f(x,y) = 1$. In this regime, a 1 qubit vs $\log(n)$ bits separation has been found for the communication complexity of the NOT-EQUAL function [122] and a 1 vs $n$ gap exists for the query complexity (the number of queries that must be made to the input bits) [55]. Furthermore, [122] uses its separation to show that unbounded classical communication is needed to simulate bipartite measurements on a Bell state if the parties share only a finite amount of randomness. In a similar vein, it was shown in [71] that there exist scenarios where a qubit can be substituted only for an unbounded number of classical bits.

## 3.2 The exclusion game

The game we consider, and shall refer to as the exclusion game, involves Alice and Bob, together with a referee to mediate the task. It runs as follows. First, the referee gives Alice an $n$-bit string, $\vec{x} \in \{0,1\}^n$, with each of the $2^n$ strings being equally likely. Alice is then allowed to send a single message regarding her input to Bob. Next, the referee chooses at random a subset, $y \subseteq [n]$ of size $m$, of locations in Alice's bit string and gives this to Bob. There are $\binom{n}{m}$ possible subsets and again they are equally likely. If $\mathcal{M}_y(\vec{x})$ denotes the $m$-bit string formed by restricting $\vec{x}$ to the bits specified by $y$, Bob's task is to produce a string, $\vec{z}_y \in \{0,1\}^m$, such that $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$.

As an illustration, consider a game where $n = 3$, $m = 2$ and the inputs given to Alice and Bob are $\vec{x} = 001$ and $y = \{1,3\}$ respectively. Winning answers that Bob can give would then be $\vec{z}_y \in \{00, 10, 11\}$ as the only losing answer is $\vec{z}_y = \mathcal{M}_y(\vec{x}) = \mathcal{M}_{\{1,3\}}(001) = 01$.

## 3.3 Preliminaries and notation for communication protocols

Before proving the existence of the claimed separations in the exclusion game, we first introduce some notation, definitions and useful lemmas.

### 3.3.1 Asymptotic complexity

We will make use of the following standard Bachmann-Landau notations for describing the asymptotic behavior of functions. More formal definitions can be found, for example, in [104].

**Definition 7** (Bachmann-Landau notation)**.** *Given a function $f(n)$, if for some function $g(n)$ and for sufficiently large $n$:*

1. *$f(n) \leq kg(n)$ for some positive constant $k$, we say that $f(n) \in O(g(n))$. The function $f$ is bounded above by $g$ (up to a constant factor) asymptotically.*

2. *$f(n) \leq kg(n)$ for every positive constant $k$, we say that $f(n) \in o(g(n))$. The function $f$ is dominated by $g$ asymptotically.*

3. *$f(n) \geq kg(n)$ for some positive constant $k$, we say that $f(n) \in \Omega(g(n))$. The function $f$ is bounded below by $g$ (up to a constant factor) asymptotically.*

4. *$f(n) \geq kg(n)$ for every positive constant $k$, we say that $f(n) \in \omega(g(n))$. The function $f$ dominates $g$ asymptotically.*

5. *$k_1 g(n) \leq f(n) \leq k_2 g(n)$ for some positive constants $k_1$ and $k_2$, we say that $f(n) \in \Theta(g(n))$. The function $f$ is bounded above and below by $g$ asymptotically.*

In addition, we shall introduce the following ideas to capture what we mean by an *infinite separation* between two positive scalings $f_1(n)$ and $f_2(n)$ in the limit of $n \to \infty$. Separations are usually characterized by increasing functions, for example, a quadratic ($f_1 \in O\left(\sqrt{f_2}\right)$) or exponential ($f_1 \in O(\log f_2)$). However, there exist gaps that grow faster than any such function:

**Definition 8** (Infinite gap)**.** *Given two functions, $f_1(n)$ and $f_2(n)$, if one function tends to infinity (zero) with $n$ and the other is asymptotically non-increasing (non-decreasing), the separation between them is said to be infinite.*

One can also define:

**Definition 9** (Doubly infinite gap)**.** *Given two functions, $f_1(n)$ and $f_2(n)$, if one function tends to infinity as $n$ increases and the other tends to zero, the separation is said to be doubly infinite.*

If one function tends to infinity or zero whilst the other tends to a non-zero constant, we shall call the separation singly infinite.

Note that one could argue that such infinite separations are not as large as some (for example) exponential separations in the following sense. Consider $f_1(n) = \log n$ and $f_2(n) = n$. Then the separation between $f_1$ and $f_2$ is exponential and $|f_2 - f_1| \in \Omega(n)$. Alternatively, consider $g_1(n) = c$ (for some positive constant $c$) and $g_2(n) = \log n$. By the above definitions, the separation between $g_1$ and $g_2$ is (singly) infinite and yet $|g_2 - g_1| \in O(\log n)$. Thus $|f_2 - f_1|$ grows asymptotically faster than $|g_2 - g_1|$ despite the separation being 'merely' exponential. However, the purpose of defining these infinite separations is to capture the properties of a function mapping between two scalings rather than the behavior of the relative difference between them.[3]

### 3.3.2 Information theory

To define complexity measures for communication protocols, we require quantities from both classical and quantum information theory. Here we give the definitions of the relevant ones. For a more thorough overview, see, for example, [134].

**Definition 10** (Entropies)**.** *In classical information theory:*

- *The Shannon entropy of a classical random variable, $X$, which takes values $x \in \mathcal{X}$, each with probability $p(x)$, is given by:*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x). \tag{3.1}$$

  *Note in particular that if $X$ has support on $|\mathcal{X}|$ elements, then $H(X) \leq \log_2 |\mathcal{X}|$ with equality if and only if $X$ is uniformly distributed over $\mathcal{X}$.*

- *For two classical random variables, $X$ and $Y$, the entropy of $X$ conditioned on knowing $Y$ (the conditional entropy of $X$ given $Y$) is given by:*

$$H(X|Y) = \sum_{y \in \mathcal{Y}} p(y) H(X|Y = y), \tag{3.2}$$

[3] For communication tasks as we consider here, the complexities under consideration are always upper bounded by $n$ (as Alice can always send the entirety of her input to Bob). In the separations we derive between the various quantum and classical complexities of the exclusion game, the relevant classical quantity, call it $C(n)$ say, will scale linearly in $n$ whilst its quantum equivalent, $Q(n)$, will tend to a constant. Hence, these infinite separations also achieve the maximum possible asymptotic behavior: $|C(n) - Q(n)| \in \Omega(n)$.

*or, equivalently:*

$$H(X|Y) = H(X,Y) - H(Y). \tag{3.3}$$

*In quantum information theory:*

- *The von Neumann entropy of a quantum state, $\rho$, belonging to a Hilbert space, $\mathcal{H}$, is given by:*

$$S(\rho) = -\mathrm{Tr}\left[\rho \log_2 \rho\right]. \tag{3.4}$$

- *For a composite system with two components, $A$ and $B$, in a joint state, $\rho_{AB}$, on a product Hilbert space, $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, the conditional quantum entropy is given by:*

$$S(A|B) = S(\rho_{AB}) - S(\rho_B), \tag{3.5}$$

*with the reduced density matrix $\rho_B$ defined by $\rho_B = \mathrm{Tr}_A\left[\rho_{AB}\right]$.*

From these quantities we can define the *mutual information* between two random variables or subspaces of a quantum state. This can be regarded as capturing the amount of information the variables or subspaces share.

**Definition 11** (Mutual information)**.** *In classical information theory:*

- *The mutual information between two classical random variables, $X$ and $Y$, is given by:*

$$I_C(X:Y) = H(X) + H(Y) - H(X,Y), \tag{3.6}$$

*or equivalently, using Eq.* (3.3)*:*

$$I_C(X:Y) = H(X) - H(X|Y). \tag{3.7}$$

- *The mutual information between two classical random variables, $X$ and $Y$, conditioned on a third variable, $Z$, is given by:*

$$I_C(X:Y|Z) = H(X|Z) - H(X|Y,Z). \tag{3.8}$$

*In quantum information theory:*

- *The quantum mutual information between two components, $A$ and $B$, of a composite quantum state, $\rho_{AB}$, on a product Hilbert space, $\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B$, is given by:*

$$I_Q(A:B) = S(\rho_A) + S(\rho_B) - S(\rho_{AB}), \tag{3.9}$$

*with the reduced density matrices $\rho_A$ and $\rho_B$ defined by $\rho_A = \mathrm{Tr}_B\left[\rho_{AB}\right]$ and $\rho_B = \mathrm{Tr}_A\left[\rho_{AB}\right]$.*

- *The quantum mutual information between two components, $A$ and $B$, conditioned on a third component, $C$, of a composite quantum state $\rho_{ABC}$, on a product Hilbert space, $\mathcal{H}_{ABC} = \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$, is given by:*

$$I_Q(A:B|C) = S(\rho_{AC}) + S(\rho_{BC}) - S(\rho_C) - S(\rho_{ABC}), \tag{3.10}$$

*with the reduced density matrices defined similarly to the above. Note that it is symmetric under interchange of $A$ and $B$. Equivalently, using Eq. (3.5) it can be written as:*

$$I_Q(A:B|C) = S(A|C) - S(A|B,C). \tag{3.11}$$

### 3.3.3 Complexity measures for communication protocols and tasks

The game we consider here is based upon using only one-way communication from Alice to Bob. To assist them with their task, they may have access to additional resources such as randomness in the from of either a public or private coin or, in the quantum setting, they may share entangled states. These will influence the various measures of the value of a given protocol and the hardness of a particular task. In this section, we define the communication complexities and information costs required for formally stating the chapter's main results.

Our exclusion game is a *relational problem*, so we being by defining what is meant by a relation:

**Definition 12** (Relational problems)**.** *Given sets $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$, a relation, $f$, is a subset $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$. In a relational problem, Alice is given $x \in \mathcal{X}$ and Bob is given $y \in \mathcal{Y}$ according some joint probability distribution $\mu$. Their task is to produce a $z \in \mathcal{Z}$ that satisfies the relation, that is, $(x, y, z) \in f$. If they produce a $z$ such that $(x, y, z) \notin f$, we say that they have made an error.*

#### Communication complexity

The classical communication complexity of the protocol $\pi$, is the maximum number of bits that two players exchange in any run of the protocol where the maximization is taken over all inputs and the value of any randomness used. Similarly, the communication complexity of a quantum protocol is defined by using qubits in place of bits. For a given relation, the *best* protocol is the one that has the minimum communication complexity. In general, there may be constraints on the resources available when performing a given task and here, we shall be interested in the following quantities:

**Definition 13** (One-way, public-coin randomized, classical communication complexity)**.** *For a relation, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, let $R^{1,pub}_{\epsilon}(f)$ denote the classical communication complexity of the best one-way, public-coin randomized, classical protocol that computes $f$ with probability of error at most $\epsilon$ on all inputs. When referring specifically to the exclusion game, we will denote this by $C_{CC}(EXC_{n,m,\epsilon})$.*

The quantum analogue of this is:

**Definition 14** (One-way, quantum communication complexity)**.** *For a relation, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, let $Q^{1}_{\epsilon}(f)$ denote the quantum communication complexity of the best one-way, quantum protocol that computes $f$ with probability of error at most $\epsilon$ on all inputs. When referring specifically to the exclusion game, we will denote this by $Q_{CC}(EXC_{n,m,\epsilon})$.*

### Communication complexity with abort

To derive an infinite separation with respect to the communication complexity, we will be interested in a modification of the exclusion game where Alice is allowed to abort the game with some probability. When she does not abort, Bob will be required to output a winning answer with certainty. For classical protocols, this leads to:

**Definition 15** (One-way, public-coin randomized, classical communication complexity with abort)**.** *For a relation, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, let $R^{1,pub}_{\delta\text{-}abort}(f)$ denote the classical communication complexity of the best one-way, public-coin randomized, classical protocol such that Alice aborts with probability at most $\delta$ on all inputs and Bob calculates $f$ with zero-error when she does not abort. When referring specifically to the exclusion game, we will denote this by $C_{CC}(EXC_{n,m,\delta\text{-}abort})$.*

The quantum strategy will make use of entanglement and so we define:

**Definition 16** (One-way, entanglement assisted communication complexity with abort)**.** *For a relation, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, let $E^{1}_{\delta\text{-}abort}(f)$ denote the classical communication complexity of the best one-way, entanglement assisted protocol such that Alice aborts with probability at most $\delta$ on all inputs and Bob calculates $f$ with zero-error when she does not abort.*[4] *When referring specifically to the exclusion game, we will denote this by $E_{CC}(EXC_{n,m,\delta\text{-}abort})$.*

[4]Note that here, as we are only interested in the asymptotic scaling of the complexity, there is little difference between allowing the players classical or quantum communication. With access to unlimited entanglement, the quantum communication can be reproduced using classical communication through quantum teleportation. Hence the entanglement assisted complexity with classical communication is at most twice as large as the entanglement assisted complexity with quantum communication.

**Information cost**

The information cost of a protocol captures the amount of information that players reveal regarding their inputs. As such, it depends on the distribution that the inputs follow.[5] More formally:

**Definition 17** (Internal information cost of a protocol)**.** *Suppose $X$ and $Y$ are distributed according to some joint distribution, $\mu$. For a protocol $\pi$, let $\pi(X,Y)$ denote the messages exchanged during the protocol together with the public randomness used. The internal information cost of $\pi$ is then:*

$$IC_{\mu}(\pi) = I_{\lambda}(X : \pi(X,Y)\,|Y) + I_{\lambda}(Y : \pi(X,Y)\,|X), \tag{3.12}$$

*where if $\pi$ is classical, we use the classical conditional mutual information ($\lambda = C$) and if it is quantum, we use its quantum definition ($\lambda = Q$).*

Intuitively, the first term in Eq. (3.12) captures the amount of information that Bob gains about Alice's input, $X$, by following the protocol $\pi$. Conditioning on $Y$ accounts for any correlations that exist between $X$ and $Y$. The second term reverses the role of Alice and Bob.

If $\pi$ only involves communication from Alice to Bob, then Eq. (3.12) reduces to:

$$\mathrm{IC}_{\mu}(\pi) = I_{\lambda}(X : \pi(X,Y)\,|Y). \tag{3.13}$$

The internal information cost of a task is then defined by:

**Definition 18** (Zero-error, one-way, classical internal information cost of a task)**.** *For a relation, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ and a distribution $\mu$ on $X$ and $Y$, let $CIC_{\mu,0}^{1,pub}(f)$ denote the information cost of the best one-way, public-coin randomized, classical protocol that computes $f$ with certainty. When referring specifically to the exclusion game, we will denote this by $C_{IC}\left(EXC_{n,m,0}^{\mu}\right)$.*

*A distribution independent measure of the classical information cost of a task under such conditions is denoted by $CIC_{0}^{1,pub}(f)$ and given by:*

$$CIC_{0}^{1,pub}(f) = \max_{\mu} CIC_{\mu,0}^{1,pub}(f). \tag{3.14}$$

*When referring specifically to the exclusion game, we shall denote this by $C_{IC}(EXC_{n,m,0})$.*

[5] To see this, consider a protocol in which Alice sends her entire input to Bob. If Alice and Bob know that their inputs are perfectly correlated, then, upon receiving Alice's message, Bob learns nothing that he did not already know. On the other hand, if their inputs are not perfectly correlated, Alice's message reveals something to Bob.

Similarly, for quantum strategies:

**Definition 19** (Zero-error, one-way, quantum internal information cost of a task)**.** *For a relation, $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ and a distribution $\mu$ on $X$ and $Y$, let $QIC^{1}_{\mu,0}(f)$ denote the information cost of the best one-way, quantum protocol that computes $f$ with certainty. When referring specifically to the exclusion game, we will denote this by $Q_{IC}\left(EXC^{\mu}_{n,m,0}\right)$.*

*A distribution independent measure of the quantum information cost of a task under such conditions is denoted by $QIC^{1}_{0}(f)$ and given by:*

$$QIC^{1}_{0}(f) = \max_{\mu} QIC^{1}_{\mu,0}(f). \tag{3.15}$$

*When referring specifically to the exclusion game, we shall denote this by $Q_{IC}(EXC_{n,m,0})$.*

### 3.3.4 Properties of and relationships between complexity measures

The following properties will be useful in deriving the separations for the exclusion game. Firstly, the information cost and the communication complexity of a classical protocol are related by:

**Lemma 3.** *For a classical protocol, $\pi_C$, for any distribution, $\mu$, over the inputs $X$ and $Y$:*

$$IC_{\mu}(\pi_C) \leq CC(\pi_C), \tag{3.16}$$

*where $CC(\pi_C)$ denotes the protocol's classical communication complexity and $IC_{\mu}(\pi_C)$ its internal information cost.*

*Proof.* See, for example, [27]. Intuitively this inequality holds as one bit of communication can carry at most one bit of information. □

In the exclusion game, Alice and Bob's inputs are independent and taken from a uniform distribution. This simplifies Eq. (3.13) as follows:

**Lemma 4.** *Suppose $X$ and $Y$ are independent and uniformly distributed ($\mu = unif$) random variables taking values in the sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Then in a one-way classical protocol, $\pi_C$:*

$$IC_{unif}(\pi_C) = \log_2 |\mathcal{X}| - H(X|\pi_C). \tag{3.17}$$

*Proof.*

$$\begin{aligned}
\mathrm{IC}_{unif}(\pi_C) &= I_C(X:\pi_C|Y), \\
&= H(X|Y) - H(X|\pi_C, Y), \\
&= H(X) - H(X|\pi_C, Y), \quad \text{as } X \text{ is independent of } Y, \\
&= H(X) + H(\pi_C, Y) - H(X, Y, \pi_C), \\
&= H(X) + H(\pi_C) + H(Y) - H(X, \pi_C) - H(Y), \quad \text{as } X, \pi_C \text{ independent of } Y, \\
&= H(X) - H(X|\pi_C), \\
&= \log_2 |\mathcal{X}| - H(X|\pi_C), \quad \text{as } X \text{ is uniformly distributed.}
\end{aligned}$$

□

As the protocols under consideration are one-way, in the quantum case their information cost can be bounded as follows:

**Lemma 5.** *Suppose $X$ and $Y$ are distributed according to $\mu$. Then the information cost of a one-way quantum protocol, $\pi_Q$, can be bounded to give:*

$$IC_\mu(\pi_Q) \leq 2S(\pi_Q). \tag{3.18}$$

*Proof.*

$$\begin{aligned}
\mathrm{IC}_\mu(\pi_Q) &= I_Q(X:\pi_Q|Y), \\
&= S(\pi_Q|Y) - S(\pi_Q|X, Y).
\end{aligned}$$

Now:

$$S(\pi_Q|Y) \leq S(\pi_Q),$$

as conditioning never increases the von Neumann entropy and:

$$\begin{aligned}
S(\pi_Q|X, Y) &= S(\pi_Q, X, Y) - S(X, Y), \\
&\geq |S(\pi_Q) - S(X, Y)| - S(X, Y), \\
&\geq -S(\pi_Q),
\end{aligned}$$

where in the penultimate inequality we have used the Araki-Lieb inequality for the joint entropy [6].[6] Applying these two bounds to the information cost, we obtain:

$$\mathrm{IC}_{\mu}\left(\pi_{Q}\right) \leq 2S(\pi_{Q}),$$

as required. □

It will also be useful to relate the communication complexity of a protocol where Alice aborts with probability at most $\delta$ on any pair of inputs, with the information complexity of a protocol in which Alice aborts with probability at most $\delta$ when the inputs are sampled according to some distribution. This is captured in the following:

**Lemma 6.** *For a relation $f \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, let $\Pi^{1,pub}_{C,(\delta,\mu)}\left(f\right)$ be the set of all classical, one-way protocols with access to shared randomness such that Alice aborts with probability at most $\delta$ (where the inputs are sampled according to $\mu$) and Bob calculates $f$ with zero error when she does not abort.*

*Let $\Pi^{1,pub}_{C,(\delta,max)}\left(f\right)$ be the set of all classical, one-way protocols with access to shared randomness such that Alice aborts with probability at most $\delta$ on any pair of inputs and Bob calculates $f$ with zero error when she does not abort. Let $\pi^{*} \in \Pi^{1,pub}_{C,(\delta,max)}\left(f\right)$ be a protocol with communication cost $R^{1,pub}_{\delta\text{-}abort}\left(f\right)$ (as defined in Definition 15).*

*Then for $\delta$ such that $0 < \delta < 1$ and any distribution $\mu$:*

$$R^{1,pub}_{\delta\text{-}abort}\left(f\right) \geq IC_{\mu}\left(\pi^{*}\right) \geq \min_{\pi \in \Pi^{1,pub}_{C,(\delta,\mu)}(f)} IC_{\mu}\left(\pi\right). \tag{3.19}$$

*Proof.* The first inequality follows from Lemma 3. To see the second inequality, note that the probability that a protocol in $\Pi^{1,\mathrm{pub}}_{C,(\delta,\mathrm{max})}\left(f\right)$ aborts when $X$ and $Y$ are distributed according to $\mu$ is:

$$\sum_{x,y} p\left(\mathrm{abort}|x,y\right)\mu\left(x,y\right) \leq \sum_{x,y} \delta\mu\left(x,y\right) \leq \delta.$$

Hence, $\pi^{*} \in \Pi^{1,\mathrm{pub}}_{C,(\delta,\mathrm{max})}\left(f\right) \subseteq \Pi^{1,\mathrm{pub}}_{C,(\delta,\mu)}\left(f\right)$. □

With these in place, we can now move onto deriving separations for the exclusion game.

[6]The Araki-Lieb inequality states that given distinct quantum systems with joint state $\rho_{AB}$, then:

$$S\left(\rho_{AB}\right) \geq \left|S\left(\rho_{A}\right) - S\left(\rho_{B}\right)\right|.$$

## 3.4 Separations from the exclusion game

### 3.4.1 Quantum information cost vs classical information cost

We begin by showing the existence of a doubly infinite separation between the quantum and classical information costs. More specifically, we shall see that there are parametrizations of $m$ in terms of $n$ such that there exists a quantum strategy for the exclusion game for which the information cost vanishes as $n$ tends to infinity. On the other hand, in the same regime, any classical strategy that succeeds with certainty must have information cost growing linearly in $n$.

The quantum strategy is based upon the PBR exclusion measurement discussed in Section 2.5.1. It leads to the following result:

**Theorem 7.** *Suppose $m \in \omega\left(n^{\frac{1}{2}+\beta}\right)$, for some constant $\beta > 0$. Then:*

$$\lim_{n\to\infty} Q_{IC}\left(EXC_{n,m,0}\right) = 0. \tag{3.20}$$

*More informally, for such $m$, there exists a quantum strategy for the exclusion game (for all prior distributions on $\vec{x}$ and $y$) such that Bob is able to produce $\vec{z}_y \neq \mathcal{M}_y\left(\vec{x}\right)$, for any $y$, while the amount of information Alice reveals to Bob regarding $\vec{x}$ tends to zero in the limit of large $n$.*

*Proof.* We first give the protocol that wins the exclusion game with certainty and then show that its information cost tends to zero in the limit of large $n$ for the specified $m$.

The protocol runs as follows:

1. Alice receives input $\vec{x} \in \{0,1\}^n$ from the referee.

2. Alice prepares the state:
$$\left|\Psi_{\vec{x}}\left(\theta_m\right)\right\rangle = \bigotimes_{i=1}^{n} \left|\psi_{x_i}\left(\theta_m\right)\right\rangle, \tag{3.21}$$
where:
$$\begin{aligned} \left|\psi_0\left(\theta\right)\right\rangle &= \cos\left(\frac{\theta}{2}\right)\left|0\right\rangle + \sin\left(\frac{\theta}{2}\right)\left|1\right\rangle, \\ \left|\psi_1\left(\theta\right)\right\rangle &= \cos\left(\frac{\theta}{2}\right)\left|0\right\rangle - \sin\left(\frac{\theta}{2}\right)\left|1\right\rangle, \end{aligned} \tag{3.22}$$
and:
$$\theta_m = 2\arctan\left(2^{1/m} - 1\right). \tag{3.23}$$

3. Alice sends $|\Psi_{\vec{x}}(\theta_m)\rangle$ to Bob.

4. Bob receives input $y$ from the referee and considers the systems in Alice's message specified by $y$. On these systems, Bob has the state:

$$|\Psi_{\mathcal{M}_y(\vec{x})}(\theta_m)\rangle = \bigotimes_{i \in y} |\psi_{x_i}(\theta_m)\rangle. \tag{3.24}$$

5. On the systems specified by $y$, Bob measures with $\mathcal{M} = \{|\zeta_{\vec{z}_y}\rangle\langle\zeta_{\vec{z}_y}|\}_{\vec{z}_y \in \{0,1\}^m}$, where:

$$|\zeta_{\vec{z}_y}\rangle = \frac{1}{\sqrt{2^m}} \left( |\vec{0}\rangle - \sum_{\vec{s} \neq \vec{0}} (-1)^{\vec{z}_y \cdot \vec{s}} |\vec{s}\rangle \right), \tag{3.25}$$

and obtains outcome $\vec{z}_y$.

6. Bob outputs $\vec{z}_y$ as the answer to the referee's question.

Note that this is a winning strategy as, from Theorem 5, $\langle \zeta_{\mathcal{M}_y(\vec{x})} | \Psi_{\mathcal{M}_y(\vec{x})}(\theta_m) \rangle = 0$. Hence, Bob always outputs $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$ and the players always succeed at their task.

To upper bound the amount of information this strategy reveals, by Lemma 5 it suffices to consider the entropy of the message sent by Alice. Furthermore, for the above strategy, the entropy of this message is maximized when the prior distribution on $\vec{x}$ and $y$ is uniform and product. To see this, using Eq. (2.42), consider the matrix of $|\Psi_{\vec{x}}\rangle\langle\Psi_{\vec{x}}|$ written in the computational basis:

$$|\Psi_{\vec{x}}\rangle\langle\Psi_{\vec{x}}| = \sum_{\vec{r},\vec{s}} (-1)^{\vec{x}\cdot(\vec{r}+\vec{s})} \left[\cos\left(\frac{\theta}{2}\right)\right]^{2n-|\vec{r}|-|\vec{s}|} \left[\sin\left(\frac{\theta}{2}\right)\right]^{|\vec{r}|+|\vec{s}|} |\vec{r}\rangle\langle\vec{s}|,$$

and note that its diagonal entries are independent of the choice of $\vec{x}$. Hence, the diagonal entries of:

$$M_Q = \sum_{\vec{x} \in \{0,1\}^n} \mu_A(\vec{x}) |\Psi_{\vec{x}}\rangle\langle\Psi_{\vec{x}}|,$$

where $\mu_A(\vec{x})$ denotes the probability of Alice's input being $\vec{x}$, are independent of the prior distribution on the players' inputs. Furthermore, $M_Q$ is a diagonal matrix when $\mu_A(\vec{x}) = \frac{1}{2^n}$, for all $\vec{x}$. Now, the Schur-Horn theorem [152, 86] implies that the diagonal elements of $M_Q$ are *majorized* by the eigenvalues of $M_Q$. That is, if $\{\lambda_i\}_{i=1}^{2^n}$ denote the eigenvalues of $M_Q$ arranged in non-increasing order and $\{d_i\}_{i=1}^{2^n}$ denote the diagonal elements of $M_Q$ arranged similarly, then:

$$\sum_{i=1}^{k} \lambda_i \geq \sum_{i=1}^{k} d_i, \quad \forall k.$$

It is a property of the Shannon entropy (see, for example, [120, Chapter 3, Section D]) that if one probability distribution, $p$, majorizes another, $q$, then $H(p) \leq H(q)$. Applying this here implies that we can upper bound the von Neumann entropy of the message sent, regardless of the prior distribution, by the Shannon entropy of the diagonal elements of $M_Q$. This is given by the von Neumann entropy of $M_Q$ when $\mu_A(\vec{x}) = \frac{1}{2^n}$, for all $\vec{x}$.

Hence, for any prior distribution on $\vec{x}$ and $y$:

$$\begin{aligned} S(M_Q) &\leq nS\left(\frac{1}{2}|\psi_0(\theta_m)\rangle\langle\psi_0(\theta_m)| + \frac{1}{2}|\psi_1(\theta_m)\rangle\langle\psi_1(\theta_m)|\right), \\ &= n\left[-\left(\left[\cos^2\left(\frac{\theta_m}{2}\right)\right]\log_2\left[\cos^2\left(\frac{\theta_m}{2}\right)\right] + \left[\sin^2\left(\frac{\theta_m}{2}\right)\right]\log_2\left[\sin^2\left(\frac{\theta_m}{2}\right)\right]\right)\right], \\ &< n\left(\frac{\theta_m}{2}\right)^2\left(\frac{1}{\ln 2} - \log_2\left[\left(\frac{\theta_m}{2}\right)^2\right]\right), \quad \text{for small } \theta_m. \end{aligned}$$

Now consider the scaling behavior of $\theta_m$. From Eq. (3.23), we have:

$$\frac{1}{m} = \log_2\left(1 + \tan\left(\frac{\theta_m}{2}\right)\right).$$

Taking the Taylor series expansion about $\theta_m = 0$ gives:

$$\frac{1}{m} = \frac{1}{\ln 2}\frac{\theta_m}{2} - \frac{1}{\ln 4}\left(\frac{\theta_m}{2}\right)^2 + \frac{{\theta_m}^3}{4\ln 8} + O\left({\theta_m}^4\right).$$

Hence, for small $\theta_m$ we have:

$$\begin{aligned} \frac{1}{m} &= \log_2\left(1 + \tan\left(\frac{\theta_m}{2}\right)\right), \\ &< \frac{1}{\ln 2}\frac{\theta_m}{2}, \\ &< \frac{2}{\ln 2}\frac{\theta_m}{2} - \frac{2}{\ln 4}\left(\frac{\theta_m}{2}\right)^2, \\ &< \frac{2}{m}. \end{aligned}$$

Using these upper and lower bounds on $\theta_m$, we obtain:

$$S(M_Q) < \frac{n}{m^2}(2\ln 2)^2\left[\frac{1}{\ln 2} + \log_2\left(\frac{m^2}{(\ln 2)^2}\right)\right], \quad \text{for large } m.$$

Hence, provided $m \in \omega\left(n^{\frac{1}{2}+\beta}\right)$, for some constant $\beta > 0$, the entropy of the message sent by Alice and the information cost of the protocol, tend to zero in the limit of large $n$. □

How much information must Alice reveal to Bob in a classical strategy? For him to succeed with certainty, the message that Alice sends needs to allow him to produce a set of answers

such that $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$ for each possible $y$. Each of these $\binom{n}{m}$ answers allows Bob to deduce a set of $2^{n-m}$ strings not equal to $\vec{x}$, although there may be some overlap between the elements in each set. To lower bound the amount of information that is revealed, we need to find the set of answers that allows Bob to exclude the fewest possible candidates for $\vec{x}$.

Doing so leads to the following result:

**Theorem 8.** *Suppose $\vec{x}$ and $y$ are chosen independently and from the uniform distribution, $\mu = unif$. Then:*

1. *In general:*

$$C_{IC}\left(EXC^{unif}_{n,m,0}\right) \geq n - \log_2(\gamma_m), \tag{3.26}$$

   *where $\gamma_m = \sum_{i=0}^{m-1}\binom{n}{i}$.*

2. *For the following parametrizations of $m$, we find:*

   (a) *If both $m \in \omega(\sqrt{n})$ and $m \in o(n)$ hold, then $C_{IC}\left(EXC^{unif}_{n,m,0}\right) \geq n - o(n)$.*

   (b) *If $m = \alpha n$ for some constant $\alpha$, $0 < \alpha < \frac{1}{2}$, then $C_{IC}\left(EXC^{unif}_{n,m,0}\right) \in \Omega(n)$.*

*More informally, for such $m$, in any classical strategy for the exclusion game that allows Bob to produce $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$, for any $y$, the amount of information that Alice reveals to Bob regarding $\vec{x}$ scales linearly in $n$.*

*Proof.* **Part 1.** Let $\pi_C$ denote a protocol followed by Alice and Bob that allows them to win the exclusion game with certainty and let $M_C$ denote the classical message Alice sends to Bob. Since Bob has to answer correctly with probability one, we can assume that Bob's strategy is deterministic (by fixing Bob's private coins). Recall that $\pi_C$ includes the public coins of the protocol and note that Alice is allowed to use private coins.

For any winning strategy, upon receiving $M_C$ from Alice, Bob must be able to construct a correct answer, $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$ for each possible $y$. We denote this set of answers by $A_{\vec{x}} = \{\vec{z}_y\}$. Each of the $\binom{n}{m}$ elements of $A_{\vec{x}}$ allows Bob to deduce a set, $S_{\vec{z}_y}$, of $2^{n-m}$ strings not equal to Alice's input, $\vec{x}$. The set $S_{\vec{z}_y}$ consists of all $\vec{x}$ such that $\mathcal{M}_y(\vec{x}) = \vec{z}_y$. Hence, each $\vec{z}_y \in A_{\vec{x}}$ reveals some information about Alice's input to Bob, although there may be some overlap between the elements in different $S_{\vec{z}_y}$. The complete set of strings that $\pi_C$ allows Bob to rule out, is denoted by $S_{\pi_C} = \cup_y S_{\vec{z}_y}$.

Let $T_{\pi_C}$ be the set of $\vec{x}$ that the protocol $\pi_C$ does not allow Bob to rule out. Then the conditional probability, $p(\vec{x}|\pi_C)$, will be non-zero only for $\vec{x} \in T_{\pi_C}$ and hence $H(X|\pi_C) \leq \log_2 |T_{\pi_C}|$. Combining this with Lemma 4 gives:

$$IC_{unif}(\pi_C) \geq n - \log_2 |T_{\pi_C}|. \tag{3.27}$$

To lower bound the information cost of a winning protocol, we need to calculate the set of answers which allows Bob to exclude the fewest possible strings.

**Claim.** *Given $\vec{x}$, the set of winning answers, $A_{\vec{x}}$, that minimizes the size of $S_{\pi_C} = \cup_{\vec{z}_y} S_{\vec{z}_y}$, is of the form $A_{\vec{x}} = \{\vec{z}_y : \vec{z}_y = \mathcal{M}_y(\vec{a}_{\vec{x}})\}$, where $\vec{a}_{\vec{x}} \in \{0,1\}^n$ is a suitably chosen bit string such that $\mathcal{M}_y(\vec{a}_{\vec{x}}) \neq \mathcal{M}_y(\vec{x})$, for all $y$.*

*Proof.* To determine the set of answers, $A_{\vec{x}}$, which minimizes the size of $S_{\pi_C} = \cup_{\vec{z}_y} S_{\vec{z}_y}$, first:

- Label the answers $\vec{z}_{y_i}$, $1 \leq i \leq \binom{n}{m}$.
- Let $k_{ij} = |y_i \cap y_j|$ be the number of places in which answers $\vec{z}_{y_i}$ and $\vec{z}_{y_j}$ overlap (i.e. refer to the same bit in $\vec{x}$). Note that $0 \leq k_{ij} \leq m-1$.
- Similarly define $k_{ij\ldots l}$ to be the number of places where answers $\vec{z}_{y_i}, \vec{z}_{y_j}, \ldots \vec{z}_{y_l}$ overlap.
- Let $r_{ij}$ be the number of places in which answers $\vec{z}_{y_i}$ and $\vec{z}_{y_j}$ agree (i.e. assign the same value to a common location in $\vec{x}$). Note that $0 \leq r_{ij} \leq k_{ij}$.

With these definitions, we proceed as follows:

- Answer $\vec{z}_{y_1}$ excludes $2^{n-m}$ strings.
- Answer $\vec{z}_{y_2}$ excludes $2^{n-m}$ strings. Some of these strings may have already been excluded by $\vec{z}_{y_1}$ and this will occur if and only if $r_{12} = k_{12}$, i.e. the two answers give the same value for the bits they overlap on. The number of strings that have already been excluded by $\vec{z}_1$ is then $\delta_{r_{12},k_{12}} 2^{n-2m+k_{12}}$, so the number of new strings excluded by $\vec{z}_{y_2}$ is:

$$2^{n-m} - \delta_{r_{12},k_{12}} 2^{n-2m+k_{12}}.$$

- Answer $\vec{z}_{y_3}$ excludes $2^{n-m}$ strings but we need to subtract the strings excluded by ($\vec{z}_{y_1}$ and $\vec{z}_{y_3}$), ($\vec{z}_{y_2}$ and $\vec{z}_{y_3}$) and add back in the strings excluded by ($\vec{z}_{y_1}$ and $\vec{z}_{y_2}$ and $\vec{z}_{y_3}$).

The number of new strings excluded is thus given by:

$$2^{n-m} - \delta_{r_{13},k_{13}} 2^{n-2m+k_{13}} - \delta_{r_{23},k_{23}} 2^{n-2m+k_{23}} + \delta_{r_{12},k_{12}} \delta_{r_{13},k_{13}} \delta_{r_{23},k_{23}} 2^{n-3m+k_{12}+k_{13}+k_{23}-k_{123}}. \tag{3.28}$$

Here $k_{123}$ is the number of locations where $\vec{z}_{y_1}$, $\vec{z}_{y_2}$ and $\vec{z}_{y_3}$ overlap.

- This construction then needs to be continued up to answer $\vec{z}_{y_{\binom{n}{m}}}$ and the number of new strings each mask excludes summed to give the total number of strings excluded.

From this construction, we see that to minimize the number of strings excluded, one way is to choose $A_{\vec{x}}$ to be such that $r_{ij} = k_{ij}$, for all $i, j$. Note that if we had $r_{13} < k_{13}$ in Eq. (3.28), then it is not possible to exclude fewer strings. To see this, observe that for three subsets $y_1, y_2$ and $y_3$ of $[n]$, each of size $m$:

$$\begin{aligned}
m &= |y_2|, \\
&\geq |y_2 \cap (y_1 \cup y_3)\,|, \\
&= |y_1 \cap y_2| + |y_2 \cap y_3| - |y_1 \cap y_2 \cap y_3|, \\
&= k_{12} + k_{23} - k_{123}, \\
\Rightarrow \quad n - 3m + k_{12} + k_{13} + k_{23} - k_{123} &\leq n - 2m + k_{13}, \\
\Rightarrow \quad 2^{n-3m+k_{12}+k_{13}+k_{23}-k_{123}} &\leq 2^{n-2m+k_{13}}, \\
\Rightarrow \quad 2^{n-2m+k_{23}} &\leq 2^{n-2m+k_{13}} + 2^{n-2m+k_{23}} - 2^{n-3m+k_{12}+k_{13}+k_{23}-k_{123}},
\end{aligned}$$

and setting $\delta_{r_{13},k_{13}} = 0$ does not exclude fewer strings. Similar arguments show that other $\delta_{r,k}$ must be non zero.

Hence, the answers should be consistent with one another i.e. $A_{\vec{x}} = \{\vec{z}_y : \vec{z}_y = \mathcal{M}_y\,(\vec{a}_{\vec{x}})\}$ where $\vec{a}_{\vec{x}} \in \{0,1\}^n$ is some suitably chosen bit string that ensures the $\vec{z}_y$ are winning answers.

□

Without loss of generality, to calculate the number of strings that such a $A_{\vec{x}}$ will exclude, we can assume $\vec{a}_{\vec{x}}$ to be the all zero string, $\vec{0}$. Here, the $\vec{x}$ that Bob can exclude are precisely those containing $m$ or more zeros. The number of remaining possibilities is given by $\gamma_m = \sum_{i=0}^{m-1} \binom{n}{i}$. Substituting $|T_{\pi_C}| = \gamma_m$ in Eq. (3.27), gives:

$$IC_{unif}(\pi_C) \geq n - \log_2(\gamma_m)\,.$$

As this holds for any classical protocol that wins the exclusion game with certainty, Eq. (3.26) holds.

**Part 2.** Given Eq. (3.26), we wish to show how it behaves for the particular $m$ given in the statement of Part 2. To do this, the following lemma will be useful:

**Lemma 7.** *[68, Page 427] Let $n \geq 1$ and $0 < q \leq \frac{1}{2}$. Then:*

$$\sum_{i=0}^{\lfloor qn \rfloor} \binom{n}{i} \leq 2^{nH(q)}, \tag{3.29}$$

*where $H(q)$ is the binary entropy of $q$.*

**Part 2a.** Here, both $m \in \omega\left(\sqrt{n}\right)$ and $m \in o(n)$ hold. Suppose $m = n^{1-\epsilon}$ where $0 < \epsilon < \frac{1}{2}$. Then:

$$\begin{aligned}
C_{\mathrm{IC}}\left(\mathrm{EXC}^{\mathrm{unif}}_{n,m,0}\right) &\geq n - \log_2 \gamma_{n^{1-\epsilon}}, \\
&> n - \log_2 \left( \sum_{i=0}^{n^{1-\epsilon}} \binom{n}{i} \right), \\
&\geq n - \log_2 \left( 2^{nH\left(n^{-\epsilon}\right)} \right), \quad \text{using Lemma 7,} \\
&= n - nH\left(n^{-\epsilon}\right), \\
&\geq n - \log_2(e)\, n^{1-\epsilon} - \epsilon n^{1-\epsilon} \log_2(n), \quad \text{for large n.}
\end{aligned}$$

Hence, for this parametrization of $m$, $C_{\mathrm{IC}}\left(\mathrm{EXC}^{\mathrm{unif}}_{n,m,0}\right) \geq n - o(n)$.

**Part 2b.** Here $m = \alpha n$, for some constant $\alpha$ such that $0 < \alpha < \frac{1}{2}$. Then:

$$\begin{aligned}
C_{\mathrm{IC}}\left(\mathrm{EXC}^{\mathrm{unif}}_{n,m,0}\right) &\geq n - \log_2 \gamma_{\alpha n}, \\
&> n - \log_2 \left( \sum_{i=0}^{\alpha n} \binom{n}{i} \right), \\
&\geq n - \log_2 \left( 2^{nH(\alpha)} \right), \quad \text{using Lemma 7,} \\
&= n - nH(\alpha).
\end{aligned}$$

Hence, for this parametrization of $m$, $C_{\mathrm{IC}}\left(\mathrm{EXC}^{\mathrm{unif}}_{n,m,0}\right) \in \Omega(n)$. □

From Theorem 7 and Theorem 8 Part 2, we obtain a doubly infinite separation between quantum and classical mechanics. For the exclusion game, there exists a quantum strategy such that for certain choices of $m$, the amount of information Alice must reveal to Bob tends to 0 in the limit of large $n$. On the other hand, for the same scaling of $m$, all classical strategies must

reveal nearly $n$ bits of information about $\vec{x}$ to Bob. Quantum mechanics allows Alice to reveal almost nothing while classically she must reveal close to everything.

In the discussion so far, we have demanded that the players' strategy should allow Bob to always output a winning string. What impact does allowing Bob to make an error with probability at most $\epsilon$ have? The scaling given in Theorem 8 is not robust against allowing a constant error. To see this, suppose that Alice sends no information to Bob and upon receiving input $y$ from the referee he is forced to guess an answer. There are $2^m$ possible strings he can give and of these only one, that which is equal to $\mathcal{M}_y(\vec{x})$, is incorrect. Hence for $\epsilon \geq \frac{1}{2^m}$, Alice does not need to send a message to Bob and thus reveals no information regarding $\vec{x}$.

It should also be noted that the quantum strategy given in Theorem 7 requires exactly $n$ qubits to be sent from Alice to Bob, while, as the information cost lower bounds the communication cost (as noted in Lemma 3), an optimal classical strategy may require fewer bits to be sent. Is it possible to win the game while sending fewer qubits? In particular, can we achieve an infinite separation between the quantum and classical communication complexities? Note that such an outcome is not ruled out by the result of [107] as this only shows that an exponential separation is the maximum achievable for scenarios where the allowed error is bounded away from zero.

Answering these questions forms the basis of the remainder of the chapter.

### 3.4.2 Quantum communication complexity vs quantum information cost

We begin by investigating whether it is possible to have an infinite separation between the quantum and classical communication complexities. To do this, we consider the consequences of the existence of a quantum protocol that wins the exclusion game whilst sending $q$ (which may depend on $n$) qubits. Such a strategy can be simulated using classical communication but will introduce some error. The amount of classical communication required to reduce the error below some threshold value is captured in the following lemma:

**Lemma 8.** *Suppose that $Q_{CC}(EXC_{n,m,\epsilon}) = q$, i.e. there exists a quantum strategy, that allows Bob to output $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$ with error at most $\epsilon$ on each pair of inputs $(\vec{x}, y)$, in which Alice sends $q$ qubits. Then, for $\epsilon' > \epsilon$:*

$$C_{CC}\left(EXC_{n,m,\epsilon'}\right) \in O\left(\left(q - \log_2\left(\epsilon' - \epsilon\right)\right) 2^q\right), \tag{3.30}$$

*In other words, there exists a classical strategy that achieves error less than $\epsilon'$ and requires $O\left(\left(q - \log_2\left(\epsilon' - \epsilon\right)\right) 2^q\right)$ bits to be sent.*

*Proof.* Without loss of generality, suppose that upon receiving $\vec{x}$, the quantum strategy involves Alice sending the $q$-qubit pure state:[7]

$$|\Phi_{\vec{x}}^Q\rangle = \sum_{j=1}^{2^q} \left(\alpha_j + i\beta_j\right) |j\rangle.$$

Assume that upon receiving this state, Bob measures it with a POVM, $\mathcal{N}_y = \left\{M_{\vec{z}_y}\right\}_{\vec{z}_y \in \{0,1\}^m}$, and outputs $\vec{z}_y$ as the answer to the referee's question. Let $p_{\vec{z}_y}^{\vec{x},y}$ denote the probability of outcome $\vec{z}_y$ and note that:

$$p_{\mathcal{M}_y(\vec{x})}^{\vec{x},y} = \mathrm{Tr}\left[|\Phi_{\vec{x}}^Q\rangle\langle\Phi_{\vec{x}}^Q| M_{\mathcal{M}_y(\vec{x})}\right] \leq \epsilon, \quad \forall \vec{x},\ \forall y.$$

A classical strategy would be for Alice to send the values of the $2^{q+1}$ numbers specifying $|\Phi_{\vec{x}}^Q\rangle$ to Bob, each rounded to $p$ bits of precision together with a bit indicating the sign. In total she sends $(p+1)\, 2^{q+1}$ classical bits. Upon receiving this message, Bob constructs the state:

$$|\Phi_{\vec{x}}^C\rangle = \frac{1}{\nu} \sum_{j=1}^{2^q} \left(\tilde{\alpha}_j + i\tilde{\beta}_j\right) |j\rangle,$$

where $\left\{\tilde{\alpha}_j, \tilde{\beta}_j\right\}$ denotes the approximations to $\{\alpha_j, \beta_j\}$ sent by Alice and $\nu$ is a normalization factor. With this classical construction, Bob samples from the classical probability distribution:

$$\left\{\tilde{p}_{\vec{z}_y}^{\vec{x},y}\right\}_{\vec{z}_y \in \{0,1\}^m} = \mathrm{Tr}\left[|\Phi_{\vec{x}}^C\rangle\langle\Phi_{\vec{x}}^C| M_{\vec{z}_y}\right],$$

and outputs his result as the answer to the referee's question.

It can be shown [113] that:

$$\left|\tilde{p}_{\vec{z}_y}^{\vec{x},y} - p_{\vec{z}_y}^{\vec{x},y}\right| \leq 20\left(2^{q/2} 2^{-p}\right), \quad \forall \vec{x},\ \forall y,\ \forall \vec{z}_y,$$

and hence, to guarantee that Bob achieves error less than $\epsilon'$, $p$ can be chosen such that:

$$\epsilon + 20\left(2^{q/2} 2^{-p}\right) \leq \epsilon'.$$

Hence $p \in O\left(q - \log_2\left(\epsilon' - \epsilon\right)\right)$ and the number of bits that need to be sent to simulate the quantum strategy is $O\left(\left(q - \log_2\left(\epsilon' - \epsilon\right)\right) 2^q\right)$. □

[7] If the strategy involved a $q$-qubit mixed state, we can always replace it with a $2q$-qubit purification without changing the asymptotic scaling.

If we can lower bound the classical communication complexity of the exclusion game for some parametrization of $\epsilon'$, then we can use Lemma 8 to obtain a lower bound on the quantum communication complexity with allowed error $\epsilon$. We thus turn to the following question. Suppose that for each pair of inputs, $(\vec{x}, y)$, Bob is allowed to output a $\vec{z}_y$ such that $\vec{z}_y = \mathcal{M}_y(\vec{x})$ with probability less than $\epsilon'$. Then how much classical communication is required from Alice so that Bob does not err with probability more than $\epsilon'$?

A useful tool for obtaining bounds on classical communication complexities is that of *rectangle bounds* (see, for example, [108] for a more comprehensive introduction). To define these, we first introduce (for one-way protocols) *rectangles* and *$\epsilon$-monochromatic functions*:

**Definition 20** (One-way rectangles)**.** *For two sets, $\mathcal{X}$ and $\mathcal{Y}$, a one-way rectangle, $R$, is defined to be a set $S \times \mathcal{Y}$ where $S \subseteq \mathcal{X}$. For a distribution, $\mu$, over $\mathcal{X} \times \mathcal{Y}$, let $\mu_R$ be the distribution formed from $\mu$ by conditioning on $R$. Let $\mu(R)$ be the probability of event $R$ under the distribution $\mu$.*

**Definition 21** (One-way $\epsilon$-monochromatic)**.** *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. A distribution, $\lambda$, on $\mathcal{X} \times \mathcal{Y}$ is called one-way $\epsilon$-monochromatic for $f$ if there exists a function, $g : \mathcal{Y} \to \mathcal{Z}$, such that:*

$$P_{XY \sim \lambda}\left[(x, y, g(y)) \in f\right] \geq 1 - \epsilon.$$

With these definitions in place, we now define rectangle bounds as follows:

**Definition 22** (Rectangle bound)**.** *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation. For a distribution, $\mu$, on $\mathcal{X} \times \mathcal{Y}$, the one-way rectangle bound is a quantity defined by:*

$$rec_{\epsilon}^{1,\mu}(f) = \min_{R} \left\{ \log_2 \frac{1}{\mu(R)} : R \text{ is one-way rectangle and } \mu_R \text{ is one-way } \epsilon\text{-monochromatic.} \right\}.$$

*The one-way rectangle bound for $f$ is:*

$$rec_{\epsilon}^{1}(f) = \max_{\mu} rec_{\epsilon}^{1,\mu}(f).$$

*If the above maximization is restricted to product distributions, we can also define:*

$$rec_{\epsilon}^{1,[]}(f) = \max_{\mu:product} rec_{\epsilon}^{1,\mu}(f).$$

The utility of rectangle bounds to the problem at hand is given by the following result from [92]:

**Theorem 9** ([92]). *Let $f \subseteq \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ be a relation and let $\epsilon \in \left[0, \frac{1}{6}\right]$. Then:*

$$R_{\epsilon}^{1,pub}(f) = \Omega\left(rec_{\epsilon}^{1,[]}(f)\right).$$

This theorem implies the following useful characterization for the classical communication complexity of the exclusion game for non-zero error:

**Lemma 9.** *To show a lower bound of $c$ for $C_{CC}(EXC_{n,m,\epsilon})$, it is sufficient to show the following. Let $S$ be any subset of $\{0,1\}^n$ of size $2^{n-c}$. Let:*

$$A_M = \{\vec{z}_y \in \{0,1\}^m : y \text{ subset of } [n] \text{ of size } m\},$$

*be any set of answers for Bob. Then, for at least $\epsilon$-fraction of:*

$$\{(\vec{x}, y) : \vec{x} \in S, y \text{ a subset of } [n] \text{ of size } m\},$$

*$\vec{z}_y$ is an incorrect answer for $\vec{x}$.*

*Proof.* By Theorem 9 and the definition of rectangle bounds, we have:

$$\mathrm{C_{CC}}\left(\mathrm{EXC}_{n,m,\epsilon}\right) = \Omega\left(\mathrm{rec}_{\epsilon}^{1,unif}\left(\mathrm{EXC}_{n,m,\epsilon}\right)\right),$$

where *unif* is the product, uniform distribution over $\mathcal{X}$ and $\mathcal{Y}$. For $R = S \times \mathcal{Y}$, the probability of event $R$ occurring is given by:

$$unif(R) = \frac{1}{2^c}.$$

Hence, if we can not find a set of answers for Bob, $A_M$, (in the language of Definition 21, a function $g$) such that $unif_R$ (the uniform distribution conditioned on $R$) is one-way $\epsilon$-monochromatic, then:

$$\mathrm{rec}_{\epsilon}^{1,unif}\left(\mathrm{EXC}_{n,m,\epsilon}\right) > c,$$

and $\mathrm{C_{CC}}\left(\mathrm{EXC}_{n,m,\epsilon}\right) = \Omega(c)$. □

Using this lemma, we can now prove the following result:

**Theorem 10.** *Suppose $2 \leq m \leq \alpha n$ where $0 < \alpha < \frac{1}{2}$ is a constant, and $\epsilon \leq 2(n+1)^{-m}$. Then:*

$$C_{CC}(EXC_{n,m,\epsilon}) \in \Omega(n). \tag{3.31}$$

*More informally, for such $m$ and $\epsilon$, any classical strategy for the exclusion game that allows Bob to produce $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$, for any $y$, except with probability $\epsilon$, has communication complexity linear in $n$.*

*Proof.* First, define $\varepsilon' = \frac{1}{\sum_{i=0}^{m}\binom{n}{i}}$ and note that for large $n$:

$$\frac{1}{\sum_{i=0}^{m}\binom{n}{i}} \geq \frac{2}{(n+1)^m},$$

holds.[8]

Our goal is to determine how large $S$ can be taken to be in Lemma 9 subject to error $\varepsilon'$. Note, that from the proof of Theorem 8 Part 1, we know that, for any choice of $A_M$, at most $\sum_{i=0}^{m-1}\binom{n}{i}$ strings can be contained in $S$ without introducing any error. An example of when this occurs is when $A_M$ is such that $\vec{z}_y = \vec{0}$, for all $y$ and $S$ consists of all strings with strictly less than $m$ zeros. Which strings can be added into this $S$ while keeping the error below $\varepsilon'$?

There are $\binom{n}{m}$ strings such that $\mathcal{M}_y(\vec{x}) = \vec{0}$ for precisely one value of $y$. These are the strings with precisely $m$ zeros. If we define $S$ as:

$$S = \left\{\vec{x} : \vec{x} \in \{0,1\}^n, \sum_{i=1}^{n} x_i \geq n - m\right\},$$

then the fraction of $\{(\vec{x}, y) : \vec{x} \in S, y \text{ subset of } [n] \text{ of size } m\}$ such that $\vec{z}_y = \vec{0}$ is an incorrect answer for $\vec{x}$, is given by:

$$\frac{\binom{n}{m}}{\binom{n}{m}\sum_{i=0}^{m}\binom{n}{i}} = \varepsilon'.$$

As $S$ consists of the maximum number of strings that produce no error and strings that produce only one error, it is clear that this is the largest $S$ can be taken to be for error given by $\varepsilon'$. Hence, by Lemma 9:

$$\begin{aligned} \mathrm{C_{CC}}\left(\mathrm{EXC}_{n,m,\varepsilon'}\right) &\in \Omega\left(n - \log_2|S|\right), \\ &= \Omega\left(n - \log_2\left(\sum_{i=0}^{m}\binom{n}{i}\right)\right), \\ \Rightarrow \mathrm{C_{CC}}\left(\mathrm{EXC}_{n,m,\varepsilon'}\right) &\in \Omega(n), \quad \text{using Theorem 8 Part 2.} \end{aligned}$$

Finally, as $\varepsilon' \geq \frac{2}{(n+1)^m}$ for large $n$, the scaling holds for error parametrized by $\epsilon$ as given in the statement of the theorem. □

Combining Lemma 8 with Theorem 10 results in the following lower bound on the quantum communication complexity of the exclusion game:

[8] To see this, note that for large $n$:

$$2\sum_{i=0}^{m}\binom{n}{i} \leq \sum_{i=0}^{m} n^i \leq \sum_{i=0}^{m}\binom{m}{i}n^i = (n+1)^m.$$

**Theorem 11.** *Suppose* $m \in o(n)$ *and* $\epsilon \leq \frac{1}{(n+1)^m}$. *Then:*

$$Q_{CC}(EXC_{n,m,\epsilon}) \in \Omega(\log n). \tag{3.32}$$

*More informally, for such* $m$ *and* $\epsilon$*, any quantum strategy for the exclusion game that allows Bob to produce* $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$*, for any* $y$*, except with probability* $\epsilon$*, has communication complexity scaling at least as* $\log(n)$.

*Proof.* Substituting $\epsilon = (n+1)^{-m}$ and $\epsilon' = 2(n+1)^{-m}$ in Lemma 8 and using Theorem 10, we have:

$$\left(q - \log\left(\frac{1}{(n+1)^m}\right)\right) 2^q \in \Omega(n),$$

which implies that, provided $m$ is not $\Omega(n)$, we have $q \in \Omega(\log(n))$. Hence the quantum strategy that minimizes the communication complexity for such $\epsilon$ must use at least on the order of $\log n$ qubits. □

From Theorem 7 and Theorem 11 we obtain a doubly infinite separation between the quantum information cost and the quantum communication complexity of the exclusion game. For some choices of $m$ and for suitably small error, an unbounded amount of quantum communication is required to carry a vanishing amount of information. We have also ruled out the existence of a beyond exponential separation between the quantum and classical communication complexity of the exclusion game in this region. However, for $m = \alpha n$ where $0 < \alpha < \frac{1}{2}$, such a separation may still exist.

### 3.4.3 Quantum communication complexity vs classical communication complexity

While we have just seen that for the exclusion game there are limits to the extent to which a separation can exist between the quantum and classical communication complexities, must a quantum strategy send $n$ qubits? Here we shall show that it is possible to compress the quantum protocol to some extent to obtain a polynomial gap for certain choices of $m$ and $\epsilon$. Furthermore by modifying the game slightly and allowing for entanglement, a more dramatic result is possible.

**Without entanglement**

In the quantum strategy for the exclusion game given in Theorem 7, upon receiving $\vec{x}$, Alice sends the state:

$$|\Psi_{\vec{x}}\rangle = \sum_{\vec{r}\in\{0,1\}^n} (-1)^{\vec{x}\cdot\vec{r}} \left[\cos\left(\frac{\theta_m}{2}\right)\right]^{n-|\vec{r}|} \left[\sin\left(\frac{\theta_m}{2}\right)\right]^{|\vec{r}|} |\vec{r}\rangle, \tag{3.33}$$

where $\theta_m = 2\arctan\left(2^{1/m}-1\right)$.

Suppose that instead of sending $|\Psi_{\vec{x}}\rangle$, Alice compresses the message by projecting the state onto the space spanned by the computational basis vectors with at most $k$ ones. Rather than sending the state given in Eq. (3.33), she instead sends:

$$|\Psi^{(k)}_{\vec{x}}\rangle = \frac{1}{\sqrt{A_k}} \sum_{\substack{\vec{r}\in\{0,1\}^n \\ |\vec{r}|\leq k}} (-1)^{\vec{x}\cdot\vec{r}} \left[\cos\left(\frac{\theta_m}{2}\right)\right]^{n-|\vec{r}|} \left[\sin\left(\frac{\theta_m}{2}\right)\right]^{|\vec{r}|} |\vec{r}\rangle, \tag{3.34}$$

where:

$$A_k = \sum_{i=0}^{k} \binom{n}{i} \left[\cos\left(\frac{\theta_m}{2}\right)\right]^{2(n-i)} \left[\sin\left(\frac{\theta_m}{2}\right)\right]^{2i}. \tag{3.35}$$

This compression reduces the number of qubits that Alice sends to $\log\left(\sum_{i=0}^{k}\binom{n}{i}\right)$. Assuming that Bob performs the same measurement on the qubits specified by $y$ as he would without the compression:

$$|\zeta_{\vec{z}_y}\rangle = \frac{1}{\sqrt{2^m}} \left(|\vec{0}\rangle - \sum_{\vec{s}\neq\vec{0}} (-1)^{\vec{z}_y\cdot\vec{s}} |\vec{s}\rangle\right), \tag{3.36}$$

this would lead to an error, $\varepsilon_k$. If $\rho^k_{\vec{x},y} = \mathrm{Tr}_{\setminus y}\left[|\Psi^{(k)}_{\vec{x}}\rangle\langle\Psi^{(k)}_{\vec{x}}|\right]$ denotes the state sent by Alice restricted to the locations specified by $y$, then:

$$\varepsilon_k = \langle\zeta_{\mathcal{M}_y(\vec{x})}|\rho^k_{\vec{x},y}|\zeta_{\mathcal{M}_y(\vec{x})}\rangle. \tag{3.37}$$

To bound $\varepsilon_k$, we make use of the following lemma:

**Lemma 10.** *For $|\Psi_{\vec{x}}\rangle$, $|\Psi^{(k)}_{\vec{x}}\rangle$ and $\varepsilon_k$, in Eqs.* (3.33), (3.34) *and* (3.37) *respectively:*

$$\sqrt{1-\left|\langle\Psi_{\vec{x}}|\Psi^{(k)}_{\vec{x}}\rangle\right|^2} \geq \varepsilon_k. \tag{3.38}$$

*Note that $\langle\Psi_{\vec{x}}|\Psi^{(k)}_{\vec{x}}\rangle$ is independent of $\vec{x}$.*

*Proof.* Recall that the *trace distance* between two density matrices, $\rho$ and $\sigma$, is given by:

$$D(\rho,\sigma) = \frac{1}{2}\mathrm{Tr}\left[\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)}\right].$$

For pure states, $|\psi\rangle$ and $|\phi\rangle$, this reduces to:

$$D\left(|\psi\rangle, |\phi\rangle\right) = \sqrt{1 - |\langle\psi|\phi\rangle|^2}.$$

We will also need the following facts. Firstly, as the trace distance never increases under local operations, for bipartite states, $\rho_{AB}$ and $\sigma_{AB}$:

$$D\left(\rho_{AB}, \sigma_{AB}\right) \geq D\left(\rho_A, \sigma_A\right).$$

Secondly [134, Page 404]:

$$D\left(\rho, \sigma\right) = \max_P \mathrm{Tr}\left[P\left(\rho - \sigma\right)\right],$$

where the maximization is taken over all projectors $P$.

Combining these facts and noting that $\langle\zeta_{\mathcal{M}_y(\vec{x})}|\rho^n_{\vec{x},y}|\zeta_{\mathcal{M}_y(\vec{x})}\rangle = 0$ gives:

$$\begin{aligned}
\varepsilon_k &= \langle\zeta_{\mathcal{M}_y(\vec{x})}|\rho^k_{\vec{x},y}|\zeta_{\mathcal{M}_y(\vec{x})}\rangle, \\
&= \langle\zeta_{\mathcal{M}_y(\vec{x})}|\rho^k_{\vec{x},y}|\zeta_{\mathcal{M}_y(\vec{x})}\rangle - \langle\zeta_{\mathcal{M}_y(\vec{x})}|\rho^n_{\vec{x},y}|\zeta_{\mathcal{M}_y(\vec{x})}\rangle, \\
&\leq D\left(\rho^k_{\vec{x},y}, \rho^n_{\vec{x},y}\right), \\
&\leq D\left(|\Psi^{(k)}_{\vec{x}}\rangle, |\Psi_{\vec{x}}\rangle\right), \\
&= \sqrt{1 - \left|\langle\Psi_{\vec{x}}|\Psi^{(k)}_{\vec{x}}\rangle\right|^2},
\end{aligned}$$

as required. □

Lemma 10 enables us to prove the following theorem.

**Theorem 12.** *Suppose that $m \in \Theta\left(n^\alpha\right)$ where $1/2 < \alpha < 1$. Then, for $\epsilon \geq \frac{1}{(n+1)^m}$:*

$$Q_{CC}\left(EXC_{n,m,\epsilon}\right) = O\left(m^{1+\delta}\right), \tag{3.39}$$

*where $\delta$ is a small positive constant.*

*More informally, for such $m$ and $\epsilon$, there exists a quantum strategy for the exclusion game such that Bob is able to produce $\vec{z}_y \neq \mathcal{M}_y\left(\vec{x}\right)$, for any $y$, except with probability $\epsilon$, and that requires on the order of $m$ qubits to be sent.*

*Proof.* The aim is to find the scaling of $k$ that achieves the required error. We want $k$ to be

such that:

$$\begin{aligned}\frac{1}{(n+1)^m} &\geq \sqrt{1-\left|\langle\Psi_{\vec{x}}|\Psi_{\vec{x}}^{(k)}\rangle\right|^2}, \\ &= \sqrt{1-\sum_{i=0}^{k}\binom{n}{i}\left[\cos\left(\frac{\theta_m}{2}\right)\right]^{2n-2i}\left[\sin\left(\frac{\theta_m}{2}\right)\right]^{2i}}, \\ &= \sqrt{\sum_{i=k+1}^{n}\binom{n}{i}\left[\cos\left(\frac{\theta_m}{2}\right)\right]^{2n-2i}\left[\sin\left(\frac{\theta_m}{2}\right)\right]^{2i}}.\end{aligned}$$

Now:

$$\begin{aligned}\binom{n}{i} &\leq \left(\frac{ne}{i}\right)^i, \\ \cos^2\left(\frac{\theta_m}{2}\right) &\leq 1, \\ \sin^2\left(\frac{\theta_m}{2}\right) &< \frac{1}{m^2}, \quad \text{for large } m,\end{aligned}$$

so, for large $m$:

$$\begin{aligned}1-\left|\langle\Psi_{\vec{x}}|\Psi_{\vec{x}}^{(k)}\rangle\right|^2 &< \sum_{i=k+1}^{n}\left(\frac{ne}{i}\right)^i\left(\frac{1}{m}\right)^{2i}, \\ &\leq (n+1)\left(\frac{ne}{m^2k}\right)^k, \quad \text{as the } i=k+1 \text{ term decays slowest for } m=\omega\left(\sqrt{n}\right).\end{aligned}$$

For this bound to be less than $\epsilon^2 = \frac{1}{(n+1)^{2m}}$, we require:

$$\begin{aligned}\left(\frac{m^2k}{ne}\right)^k &> (n+1)^{2m+1}, \\ k\log\left(\frac{m^2k}{ne}\right) &> (2m+1)\log(n+1).\end{aligned}$$

To satisfy this asymptotically, it suffices to take $k = m^{1+\beta}$, where $\beta > 0$. The number of qubits sent (which by Lemma 10 achieves an error less than $\frac{1}{(n+1)^m}$) is then:

$$\begin{aligned}\log\left(\sum_{i=0}^{m^{1+\beta}}\binom{n}{i}\right) &\leq \log\left((n+1)^{m^{1+\beta}}\right), \\ &= m^{1+\beta}\log(n+1).\end{aligned}$$

Hence, $\mathrm{Q_{CC}}\left(\mathrm{EXC}_{n,m,\epsilon}\right) = O\left(m^{1+\delta}\right)$, where $\delta > 0$. □

Combining this with Theorem 10, we see that when $m \in \Theta\left(n^\alpha\right)$ with $\frac{1}{2} < \alpha < 1$, and when the allowed error is $(n+1)^{-m}$, it is possible to have a polynomial separation between the classical and quantum communication complexities.

**With entanglement**

By modifying the game, we can obtain a task that admits a strategy involving entanglement with constant communication complexity while all classical strategies involve at least $\Omega(n)$ bits being sent. In what follows, Alice may choose to abort the game with probability $\delta$ on each pair of inputs $(\vec{x}, y)$ and the players have access to both private and shared randomness. However, when she does not abort, Bob must give a correct answer.

How does this change affect the classical communication complexity?

**Theorem 13.** *Suppose $m = \alpha n$ where $0 < \alpha < \frac{1}{2}$ and $\delta > 0$. Then:*

$$C_{CC}(EXC_{n,m,\delta\text{-}abort}) \in \Omega(n). \tag{3.40}$$

*More informally, for such $m$, any classical strategy for the exclusion game such that Alice aborts with probability at most $\delta$ on each pair of inputs and when she does not abort, Bob is able to produce $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$, has communication complexity linear in $n$.*

*Proof.* Since Bob has to answer correctly with probability one (on a non-abort message from Alice), we can again assume that Bob's strategy is deterministic (by fixing Bob's private coins). Recall that $\pi_C$ includes the public coins of the protocol and note that Alice is allowed to use private coins.

Using Lemma 6, it suffices to show that if $\vec{x}$ and $y$ are chosen independently and from the uniform distribution, $\mu = unif$, any strategy that aborts with probability at most $\delta$ but allows Bob to answer correctly otherwise, is such that $IC_{unif}(\pi_C) \in \Omega(n)$.

Consider $H(X|\pi_C)$. Using Eq. (3.2):

$$H(X|\pi_C) = p(\text{abort})\, H(X|\text{Alice aborts}) + p(\text{non abort})\, H(X|\text{Alice does not abort}).$$

To obtain a bound on $IC_{unif}(\pi_C)$, we need to upper bound this quantity. The first conditional entropy in the sum is trivially upper bounded by $n$ as in general, $H(S|T) \leq H(S)$. If Alice does not abort, then, for any input $y$, Bob must win the game with certainty. This means that we can apply the reasoning from the proof of Part 1 of Theorem 8 to upper bound the second conditional entropy by $\log_2 \gamma_m$.

This gives:

$$\begin{aligned} IC_{unif}(\pi_C) &\geq n - (\delta n + (1-\delta)\log_2 \gamma_m), \\ &= (1-\delta)(n - \log_2 \gamma_m). \end{aligned}$$

From the proof of Part 2b of Theorem 8, we know that for $m = \alpha n$ with $0 < \alpha < \frac{1}{2}$, this expression is $\Omega(n)$. □

With access to entangled states, rather than sending $|\Psi_{\vec{x}}(\theta_m)\rangle$ to Bob directly, Alice could instead attempt to steer Bob's side of the entanglement to the desired state by performing an appropriate measurement on her own system. To see how this would work, suppose Alice and Bob share $n$ entangled states, one for each bit in $\vec{x}$. From [150] we know that there exists an entangled state, $|\Phi\rangle_{AB}$, and two measurements with outcomes labeled by 0 and 1, $\mathcal{S} = \{S_0, S_1\}$ and $\mathcal{R} = \{R_0, R_1\}$, with the following properties. Firstly, if Alice measures her half of $|\Phi\rangle_{AB}$ with $\mathcal{S}$ and obtains the outcome 0, Bob's half of $|\Phi\rangle_{AB}$ is steered to $|\psi_0(\theta_m)\rangle$ while if she obtains outcome 1, Bob's system is steered to the state $|-\rangle$. Similarly, measuring with $\mathcal{R}$ will steer Bob to either $|\psi_1(\theta_m)\rangle$ or $|+\rangle$. If the value of $x_i$ determines which of $\mathcal{S}$ and $\mathcal{R}$ Alice applies, the probability that Bob's system is steered to the state $|\psi_{x_i}(\theta_m)\rangle$ is [150]:

$$p_{\text{steer}} = \frac{1}{1 + \sin\theta_m}. \tag{3.41}$$

Making use of this steering while allowing Alice to occasionally abort gives the following result:

**Theorem 14.** *Suppose $m = \alpha n$ where $0 < \alpha < \frac{1}{2}$ and $\delta > 0$. Then:*

$$E_{CC}(EXC_{n,m,\delta\text{-}abort}) \leq \log_2 k. \tag{3.42}$$

*Here $k$ is some constant that depends on $\delta$ but not on $n$.*

*More informally, for such $m$, there exists an entanglement assisted strategy for the exclusion game using $\log_2 k$ bits of communication, such that Alice aborts with probability at most $\delta$ on each pair of inputs and when she does not abort, Bob is able to produce $\vec{z}_y \neq \mathcal{M}_y(\vec{x})$.*

*Proof.* Making use of the state targeting strategy of [150], suppose Alice and Bob share $k$ sets of $n$ copies of $|\Phi_{AB}\rangle$ where:

$$|\Phi_{AB}\rangle = \sqrt{\frac{1}{2}\left(1 + \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|00\rangle + \sqrt{\frac{1}{2}\left(1 - \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|11\rangle. \tag{3.43}$$

Figure 3.1 shows the reduced state of $|\Phi\rangle_{AB}$ on Bob's Bloch sphere. This reduced state is given by:

$$\rho_B = \begin{pmatrix} \frac{1}{2}\left(1 + \frac{\cos\theta_m}{1+\sin\theta_m}\right) & 0 \\ 0 & \frac{1}{2}\left(1 - \frac{\cos\theta_m}{1+\sin\theta_m}\right) \end{pmatrix}. \tag{3.44}$$

The properties of the following measurement will be useful for our protocol.

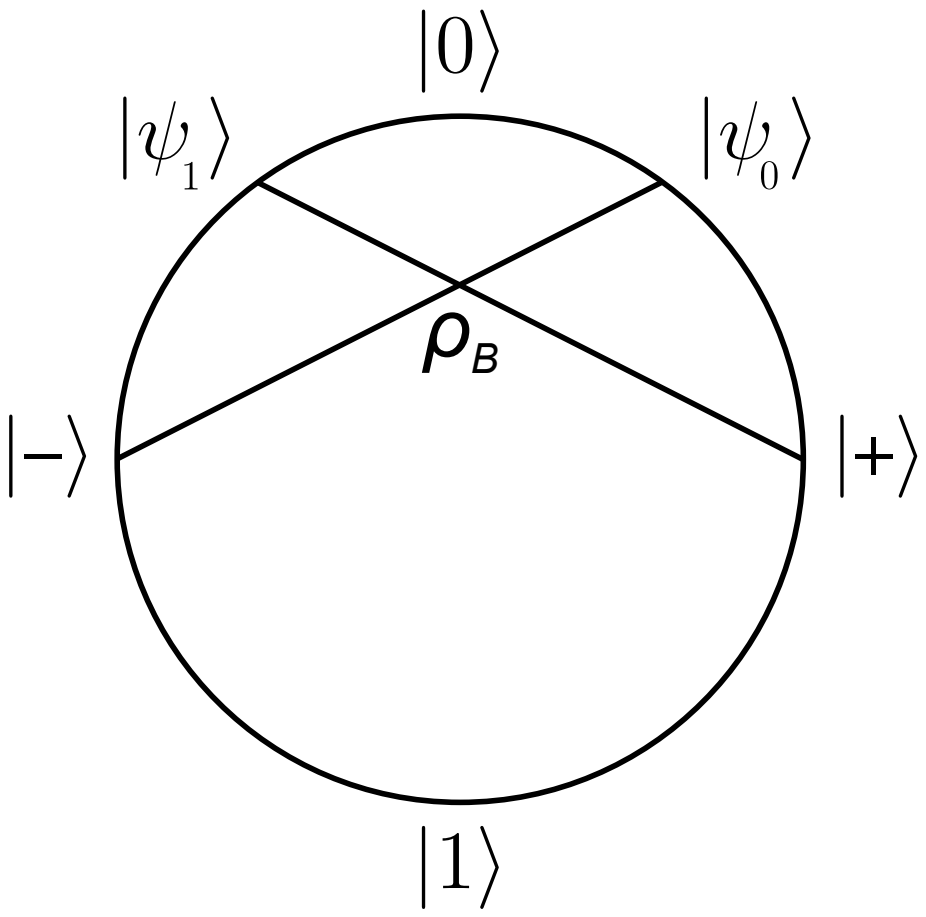



Figure 3.1: *The entangled state for the exclusion game steering strategy.* Here we show $|\Phi\rangle_{AB}$ as viewed on Bob's Bloch sphere [150]. His reduced state is denoted by $\rho_B$.

**Claim.** *Suppose Alice measures the state* $|\Phi\rangle_{AB}$ *with:*

1. $\mathcal{S} = \{S_0, S_1\}$, *where:*

$$\begin{aligned} S_0 &= |s_0\rangle\langle s_0|, \text{ with } |s_0\rangle = \sqrt{\frac{1}{2}\left(1 + \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|0\rangle + \sqrt{\frac{1}{2}\left(1 - \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|1\rangle, \\ S_1 &= |s_1\rangle\langle s_1|, \text{ with } |s_1\rangle = \sqrt{\frac{1}{2}\left(1 - \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|0\rangle - \sqrt{\frac{1}{2}\left(1 + \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|1\rangle. \end{aligned} \tag{3.45}$$

   *Then, if the outcome labeled 0 occurs, Bob's half of* $|\Phi\rangle_{AB}$ *is steered to* $|\psi_0(\theta_m)\rangle$*. This happens with probability* $\frac{1}{1+\sin\theta_m}$*. If the outcome labeled 1 occurs, Bob's half of* $|\Phi\rangle_{AB}$ *is steered to* $|-\rangle$*. This happens with probability* $\frac{\sin\theta_m}{1+\sin\theta_m}$*.*

2. $\mathcal{R} = \{R_0, R_1\}$, *where:*

$$\begin{aligned} R_0 &= |r_0\rangle\langle r_0|, \text{ with } |r_0\rangle = \sqrt{\frac{1}{2}\left(1 + \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|0\rangle - \sqrt{\frac{1}{2}\left(1 - \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|1\rangle, \\ R_1 &= |r_1\rangle\langle r_1|, \text{ with } |r_1\rangle = \sqrt{\frac{1}{2}\left(1 - \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|0\rangle + \sqrt{\frac{1}{2}\left(1 + \frac{\cos\theta_m}{1+\sin\theta_m}\right)}|1\rangle. \end{aligned} \tag{3.46}$$

   *Then, if the outcome labeled 0 occurs, Bob's half of* $|\Phi\rangle_{AB}$ *is steered to* $|\psi_1(\theta_m)\rangle$*. This happens with probability* $\frac{1}{1+\sin\theta_m}$*. If the outcome labeled 1 occurs, Bob's half of* $|\Phi\rangle_{AB}$ *is steered to* $|+\rangle$*. This happens with probability* $\frac{\sin\theta_m}{1+\sin\theta_m}$*.*

*Proof.* We give the proof for the measurement $\mathcal{S}$, the result for $\mathcal{R}$ follows similarly. The probabilities for each of the measurement outcomes can be calculated from [150].

Given a normalized entangled state, $a_0|00\rangle + a_1|11\rangle$, suppose Alice performs a measurement on her half and obtains the outcome associated with the projector $x_0|0\rangle + x_1|1\rangle$ with probability $p$. It is easy to see that Bob's system is steered to the (normalized) state:

$$\frac{1}{\sqrt{p}}\left(x_0 a_0|0\rangle + x_1 a_1|1\rangle\right). \tag{3.47}$$

Using this together with Eq. (3.43), to prove Part 1 it suffices to show that:

- For the projective measurement $|s_0\rangle$:

$$\begin{aligned}\sqrt{1+\sin\theta_m}\frac{1}{2}\left(1+\frac{\cos\theta_m}{1+\sin\theta_m}\right) &= \frac{1}{2}\sqrt{\frac{\left(1+\sin\theta_m+\cos\theta_m\right)^2}{1+\sin\theta_m}},\\ &= \frac{1}{2}\sqrt{\frac{2+2\sin\theta_m+2\cos\theta_m+2\sin\theta_m\cos\theta_m}{1+\sin\theta_m}},\\ &= \frac{1}{2}\sqrt{\frac{\left(1+\sin\theta_m\right)\left(2+2\cos\theta_m\right)}{1+\sin\theta_m}},\\ &= \cos\left(\frac{\theta_m}{2}\right),\end{aligned}$$

and similarly:

$$\begin{aligned}\sqrt{1+\sin\theta_m}\frac{1}{2}\left(1-\frac{\cos\theta_m}{1+\sin\theta_m}\right) &= \frac{1}{2}\sqrt{\frac{\left(1+\sin\theta_m-\cos\theta_m\right)^2}{1+\sin\theta_m}},\\ &= \frac{1}{2}\sqrt{\frac{2+2\sin\theta_m-2\cos\theta_m-2\sin\theta_m\cos\theta_m}{1+\sin\theta_m}},\\ &= \frac{1}{2}\sqrt{\frac{\left(1+\sin\theta_m\right)\left(2-2\cos\theta_m\right)}{1+\sin\theta_m}},\\ &= \sin\left(\frac{\theta_m}{2}\right),\end{aligned}$$

so Bob is steered to $|\psi_0\left(\theta_m\right)\rangle$ if measurement $\mathcal{S}$ is performed and outcome 0 occurs.

- For the projective measurement $|s_1\rangle$:

$$\begin{aligned}\sqrt{\frac{1+\sin\theta_m}{\sin\theta_m}}\frac{1}{2}\sqrt{1+\frac{\cos\theta_m}{1+\sin\theta_m}}\sqrt{1-\frac{\cos\theta_m}{1+\sin\theta_m}} &= \frac{1}{2}\sqrt{\frac{1+\sin\theta_m}{\sin\theta_m}}\sqrt{1-\frac{\cos^2\theta_m}{\left(1+\sin\theta_m\right)^2}},\\ &= \frac{1}{2}\sqrt{\frac{1+2\sin\theta_m+\sin^2\theta_m-\cos^2\theta_m}{\sin\theta_m+\sin^2\theta_m}},\\ &= \frac{1}{\sqrt{2}},\end{aligned}$$

so Bob is steered to $|-\rangle$ if measurement $\mathcal{S}$ is performed and outcome 1 occurs.

□

We now give an explicit protocol using these sets of $|\Phi\rangle_{AB}$, and based on the strategy in [150], that requires $\log_2 k$ bits of classical communication:

1. Alice receives $\vec{x}$ from the referee.

2. For each of the $k$ sets, on the $i^{th}$ copy of $|\Phi\rangle_{AB}$ in that set:

   (a) If $x_i = 0$, Alice measures with $\mathcal{S} = \{S_0, S_1\}$. If the outcome labeled 0 occurs, Bob's half of $|\Phi\rangle_{AB}$ is steered to $|\psi_0(\theta_m)\rangle$. If the outcome labeled 1 occurs, Bob's half of $|\Phi\rangle_{AB}$ is steered to $|-\rangle$.

   (b) If $x_i = 1$, Alice measures with $\mathcal{R} = \{R_0, R_1\}$. If the outcome labeled 0 occurs, Bob's half of $|\Phi\rangle_{AB}$ is steered to $|\psi_1(\theta_m)\rangle$. If the outcome labeled 1 occurs, Bob's half of $|\Phi\rangle_{AB}$ is steered to $|+\rangle$.

3. If there is a set in which all of the measurements resulted in the 0 outcome, Alice sends a classical message of length $\log_2 k$ to Bob indicating which set it was. Otherwise, in each of the $k$ sets, the measurement outcome 1 occurs at least once so Alice aborts the game and sends a special 'abort' symbol to Bob.

4. If Alice did not abort, Bob now has a set of $n$ states that he knows is in the state, $|\Psi_{\vec{x}}(\theta_m)\rangle$. He runs steps 4-6 of the original quantum protocol given in Theorem 7 on this set to output a winning answer.

This protocol uses $\log_2 k$ bits of communication and allows Bob to always output a winning answer when Alice does not abort. It remains to show that $k$ can be chosen to be a constant if Alice is allowed to abort with probability $\delta$.

**Claim.** *For $m = \alpha n$, the probability that Alice aborts in the above strategy, $p_{abort}$, is such that:*

$$p_{abort} \leq \left(1 - 4^{-\frac{1}{\alpha}}\right)^k. \tag{3.48}$$

*Proof.* For each measurement, the probability that Alice obtains the outcome 0 is given by [150]:

$$p_{\text{steer}} = \frac{1}{1 + \sin\theta_m}. \tag{3.49}$$

The probability that all $n$ measurements in a set give outcome 0 is then:

$$\begin{aligned} p_{\text{steer}}^{\text{global}} &= \left(\frac{1}{1+\sin\theta_m}\right)^n, \\ &= \left(1 + 2^{\frac{m-2}{m}} - 2^{\frac{m-1}{m}}\right)^n, \end{aligned}$$

where we have used the fact that $\theta_m = 2\arctan\left(2^{1/m} - 1\right)$ and the identity $\sin(2\arctan x) = \frac{2x}{1+x^2}$.

Now, setting $m = \alpha n$:

$$\begin{aligned}
\lim_{n\to\infty} p_{\text{steer}}^{\text{global}} &= \lim_{n\to\infty} \left(1 + 2^{\frac{\alpha n-2}{\alpha n}} - 2^{\frac{\alpha n-1}{\alpha n}}\right)^n, \\
&= \lim_{n\to\infty} \exp\left[n \ln\left[1 + 2^{\frac{\alpha n-2}{\alpha n}} - 2^{\frac{\alpha n-1}{\alpha n}}\right]\right], \\
&= \exp \lim_{t\to 0} \frac{\ln\left[1 + 2^{1-\frac{2t}{\alpha}} - 2^{1-\frac{t}{\alpha}}\right]}{t}, \\
&= \exp \lim_{t\to 0} \frac{\frac{2}{\alpha}\ln 2\left(-2^{1-\frac{2t}{\alpha}}\right) - \frac{1}{\alpha}\ln 2\left(-2^{1-\frac{t}{\alpha}}\right)}{1 + 2^{1-\frac{2t}{\alpha}} - 2^{1-\frac{t}{\alpha}}}, \quad \text{using l'Hopital's rule,} \\
&= \exp\left[\frac{2}{\alpha}\ln 2 - \frac{4}{\alpha}\ln 2\right], \\
&= 4^{-\frac{1}{\alpha}}.
\end{aligned}$$

As $p_{\text{steer}}^{\text{global}}$ is monotonically decreasing in $n$, $p_{\text{steer}}^{\text{global}} \geq 4^{-\frac{1}{\alpha}}$.

Finally, Alice aborts if each of the $k$ sets fail to steer globally so:

$$p_{\text{abort}} = \left(1 - p_{\text{steer}}^{\text{global}}\right)^k \leq \left(1 - 4^{-\frac{1}{\alpha}}\right)^k. \tag{3.50}$$

□

Hence, by choosing $k$ such that $\left(1 - 4^{-\frac{1}{\alpha}}\right)^k \leq \delta$, Alice and Bob succeed through sending a constant amount of classical communication, regardless of the value of $n$. □

From Theorems 13 and 14 we obtain a singly infinite separation. By allowing Alice to occasionally decline to answer, there exist choices of $m$ such that in the exclusion game, with access to entanglement, only a constant amount of communication is required. For classical strategies on the other hand, Alice still needs to send $\Omega(n)$ bits of communication. Interestingly this separation occurs for precisely the scaling of $m$ for which a beyond exponential separation was not ruled out in Section 3.4.2.

## 3.5 Summary

In this chapter we have designed a communication task that exploits a result from the foundations of quantum mechanics, the PBR theorem. Quantum strategies for this task can drastically outperform classical ones with respect to the amount of information they reveal. Additionally, when Alice is allowed to abort with some probability, the communication complexity is similarly improved by using shared entanglement. This contrasts sharply with the usual measure studied

in communication tasks, the communication complexity, where, in the absence of entanglement and for bounded error, at most an exponential advantage can be gained. Furthermore, the task also exhibits a separation between the required information and communication cost of quantum strategies which is not known to hold classically. An increasing number of qubits may be needed to send a vanishing amount of information. Figure 3.2 summarizes the results of this chapter.

It remains an open question to fully characterize how the various complexities scale for general error. In particular, is it possible to obtain a beyond exponential separation between the communication complexities without using entanglement?

Furthermore, as noted at the end of Section 3.4.1, the derived separations for the exclusion game are not at all robust to error. An interesting question is whether it is possible to find a variation on the exclusion game that alleviates this sensitivity. Constructing such a game, which still exhibits the entanglement assisted infinite separation shown here, could have striking implications regarding the achievable ratio between quantum and classical values of Bell inequalities [33]. One could envisage achieving this by requiring that Bob exclude more than one string in the exclusion game. The results of [60] may provide insight into constructing PBR-like quantum strategies for such modifications.

Finally, what does the existence of these infinite separations tell us about the structure and power of quantum mechanics? These results imply that even though a quantum message may convey a vanishingly small amount of information, to reproduce this information using purely classical means can require an infinitely large amount of information to be sent. The amount of excess informational baggage that a classical model of quantum theory needs to carry round can be very heavy indeed.

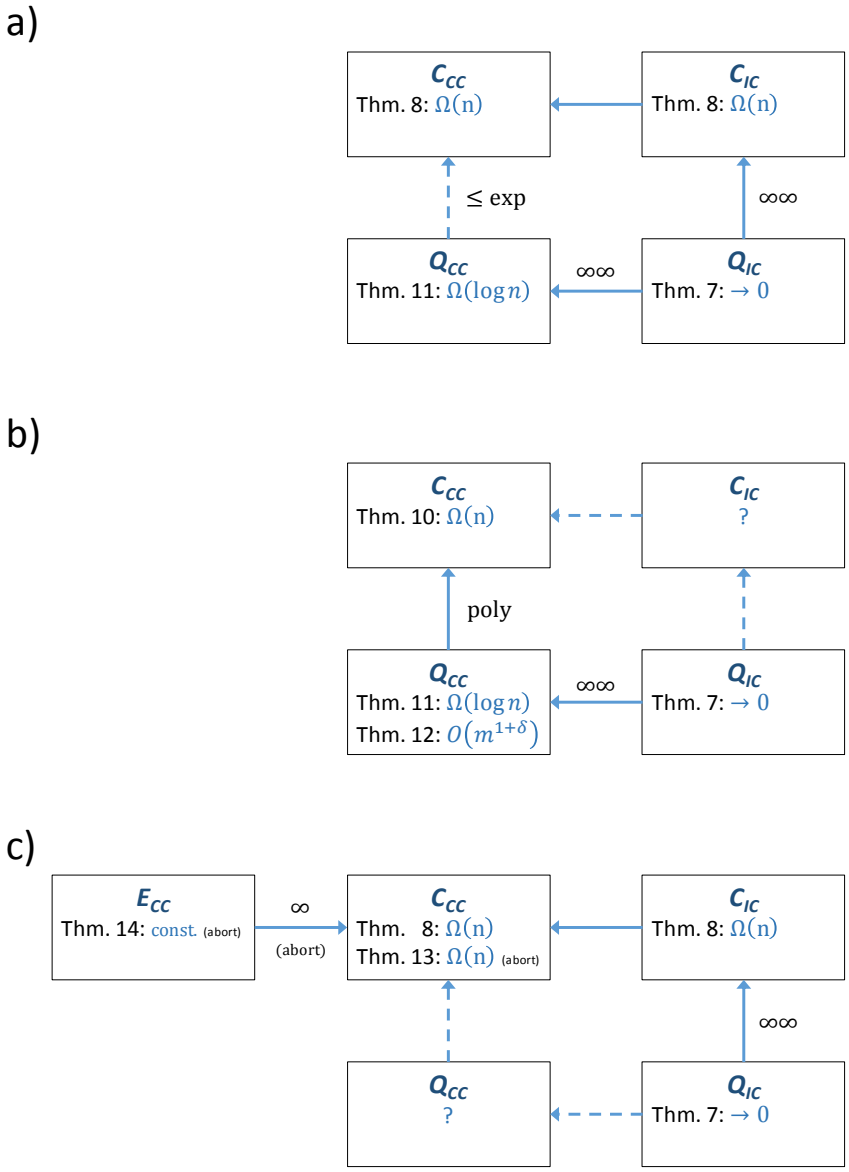



Figure 3.2: *Summary of separations for the exclusion game.* Complexities of $\mathrm{EXC}_{n,m,\epsilon}$ for: *a)* $m$ such that both $m \in \omega\left(n^{\frac{1}{2}+\beta}\right)$ where $\beta > 0$, and $m \in o(n)$ hold, with $\epsilon = 0$. *b)* $m$ such that both $m \in \omega\left(n^{\frac{1}{2}+\beta}\right)$ where $\beta > 0$, and $m \in o(n)$ hold, with $\epsilon = (n+1)^{-m}$. *c)* $m = \alpha n$ where $0 < \alpha < \frac{1}{2}$, with $\epsilon = 0$. Solid arrows denote established gaps (pointing towards the larger complexity), while the dashed ones denote unknown gaps.

# Part II

# Beyond the Thermodynamic Limit

# Guide to Part II

We now turn to the thermodynamics of small systems. Here, the framework of resource theories has proven a useful tool for analyzing the feasibility of transitioning from one state to another when we do not take the thermodynamical limit. Chapter 4 reviews two such resource theories: noisy operations [88] and thermal operations [96, 89]. Within thermal operations, the concept of thermo-majorization (generalizing that of majorization from entanglement theory) provides necessary and sufficient criteria for determining whether a given thermodynamical transition is possible at the nano-scale. Furthermore, thermo-majorization supplies us with a valuable implement for visualizing transformations and for calculating the work cost/yield of such processes.

Using thermo-majorization to address the question: 'What is the probability of a thermodynamical transition?', is the subject of Chapter 5. In the absence of coherences in the energy eigenbasis, for any two given states it is possible to make a transition provided enough work is supplied. Without this additional work, can the transformation be performed probabilistically? By adapting results from pure state entanglement theory, in this chapter we characterize this probability and give a geometric interpretation for its calculation through thermo-majorization diagrams. These diagrams also provide an efficient method for deriving the work cost/yield of a state conversion by considering only a finite set of thermodynamical monotones and allow this quantity to be related to the aforementioned probability. These characterizations were originally found in joint work with Álvaro Alhambra and Jonathan Oppenheim [4].

Thermal operations assume that one has the ability to precisely manipulate all of the degrees of freedom in an extremely large heat bath. As such, while they are useful for deriving ultimate limits and constraints on the thermodynamics of small, closed systems, they are not regarded as experimentally tractable. In Chapter 6, we define a more experimentally friendly set of operations, termed coarse operations. Surprisingly, these show that, in the absence of

coherences, one need only have control over a single qubit in the bath to implement a transition allowed by thermal operations. These results appear in [139], joint work with Piotr Ćwikliński, Janet Anders, Michał Horodecki and Jonathan Oppenheim.

# Chapter 4

# Thermodynamics as a Resource Theory

## 4.1 Thermodynamics without the thermodynamic limit

The field of thermodynamics traditionally concerns itself with the physics of large, classical systems. Here the traditional four laws of thermodynamics hold sway, providing constraints that the energy, temperature and entropy of a system must obey. With thermodynamics, one can analyze the efficiency of heat engines, the plausibility of a chemical reaction and the physics of black holes [19, 13].

Thermodynamics is particularly suited towards answering questions about what can be achieved given a particular set of resources. For example, given a system in a state, $\rho$, with some Hamiltonian, $H_1$, when can it be deterministically transformed into another state, $\sigma$, associated with a potentially different Hamiltonian, $H_2$? If our only resource is a heat bath at temperature $T$, then in the thermodynamic limit where the system is composed of many particles and if the interactions are short ranged, the answer is given in terms of the *Helmholtz free energy*. For the system $(\rho, H_1)$, this is defined by:

$$F(\rho, H_1) = \mathrm{Tr}[H_1\rho] - k_B T S(\rho), \tag{4.1}$$

where here, $S(\rho) = -\mathrm{Tr}[\rho \ln \rho]$ is the entropy of the system and $k_B$ is Boltzmann's constant. Under the above conditions, the transition from $(\rho, H_1)$ to $(\sigma, H_2)$ is possible if and only if:

$$F(\rho, H_1) \geq F(\sigma, H_2), \tag{4.2}$$

and this can be regarded as a formulation of the second law of thermodynamics, if we account for the first law of energy conservation. Moreover, in this limit, the maximum amount of work that can be extracted or must be supplied in converting $(\rho, H_1)$ into $(\sigma, H_2)$ is given by $F(\rho, H_1) - F(\sigma, H_2)$.

However, what if we are interested in small, finite systems or in systems with long-range interactions? The thermodynamics of such systems, where the thermodynamic limit does not apply, is becoming increasingly relevant as systems are cooled and manipulated at the nano-scale. For instance, as a flavor of what has been achieved, heat engines can be constructed from masers [153, 154], thermodynamical ratchets have been realized using both optics [66] and molecules [155], quantum systems have been algorithmically cooled [18] and fluctuation theorems ([97, 51]) have been experimentally tested using both RNA [49] and DNA [127, 118]. In this regime, Eq. (4.2) is no longer sufficient for determining whether a transition from $(\rho, H_1)$ to $(\sigma, H_2)$ is achievable [89] and to investigate what is possible, concepts from information theory have recently come to the fore.

The interplay between information theory and thermodynamics is perhaps best illustrated by Szilard's engine [161] which converts pure states (maximal information about a system's state) into work in the presence of a heat bath at temperature $T$. This engine consists of a single molecule gas in an isothermal box that is in contact with the heat bath together with a work storage system, such as a weight that can be connected to the system using a piston and a series of suitable pulleys. Initially, the single molecule system is in thermal equilibrium and equally likely to be on either side of the box. The protocol to extract work then runs as follows (see also Figure 4.1):

1. A partition is inserted at the center of the box, dividing it into two boxes, one on the left, $L$, and one on the right, $R$.

2. A measurement is performed to determine whether the particle is in $L$ or $R$.

3. Depending on the measurement outcome:

   (a) If the particle is in $L$, the partition is removed and a moveable piston inserted in its place such that the weight will be raised if the piston moves to the right.

   (b) If the particle is in $R$, the partition is removed and a moveable piston inserted in its place such that the weight will be raised if the piston moves to the left.

4. The gas is allowed to expand quasi-statically and isothermally causing the piston to move, the weight to be raised and the system to return to its original state.

In this process, all steps are in principle work-neutral with the exception of Step 4. Here the ideal gas law, $pV = k_B T$ (where $p$ is the pressure and $V$ is the volume), applies and the work extracted, $W_{\text{ext}}$, is given by:

$$W_{\text{ext}} = \int_{V_0/2}^{V_0} p \, \mathrm{d}V = \int_{V_0/2}^{V_0} \frac{k_B T}{V} \, \mathrm{d}V = k_B T \ln 2, \tag{4.3}$$

where $V_0$ is the initial volume of the gas. Hence, 1 bit of information, the knowledge of whether the particle is on the left or right of the box, has been converted into $k_B T \ln 2$ units of work.

At first glance, one may wonder why the above protocol does not form a perpetual motion machine: thermal energy has been converted into work and the system returned to its original state in a seemingly cyclic process. However, the process is not quite cyclic. In performing the measurement in Step 2, the experimenter has had to record the outcome using a single bit of memory, which can without loss of generality be assumed to have been initialized as $L$. After Step 4, this memory is still in the post measurement state and has not been reset to its original value - the procedure is not cyclic. To complete the cycle, the memory needs to be erased and, by Landauer's principle [109, 20], this costs $k_B T \ln 2$ units of work in the presence of a heat bath at temperature $T$ (assuming that the Hamiltonian associated with the memory is degenerate). Hence the perpetual motion is avoided.

In this thesis, the main information theoretic tool for analyzing the possibilities of thermodynamical transitions at the nano-scale will be that of *resource theories*. In particular, the resource theories of *noisy operations* and *thermal operations* will be of great importance and we define these in the next section.

## 4.2 Resource theories

Quantum information theory is often concerned with determining what it is possible to achieve given access only to a restricted set of operations, $\mathcal{C}$. This has led to the notion of quantum resource theories [90]. Under $\mathcal{C}$, it will often be the case that there exists a set of states, $\mathcal{S}$, that can always be created. These states are referred to as the *free states* of the theory as, being always producible, they are not of much value. The states that do not belong to $\mathcal{S}$ are regarded as *resource states* and having access to them may enable one to produce other states outside of $\mathcal{S}$ or to perform operations outside of the restricted set.

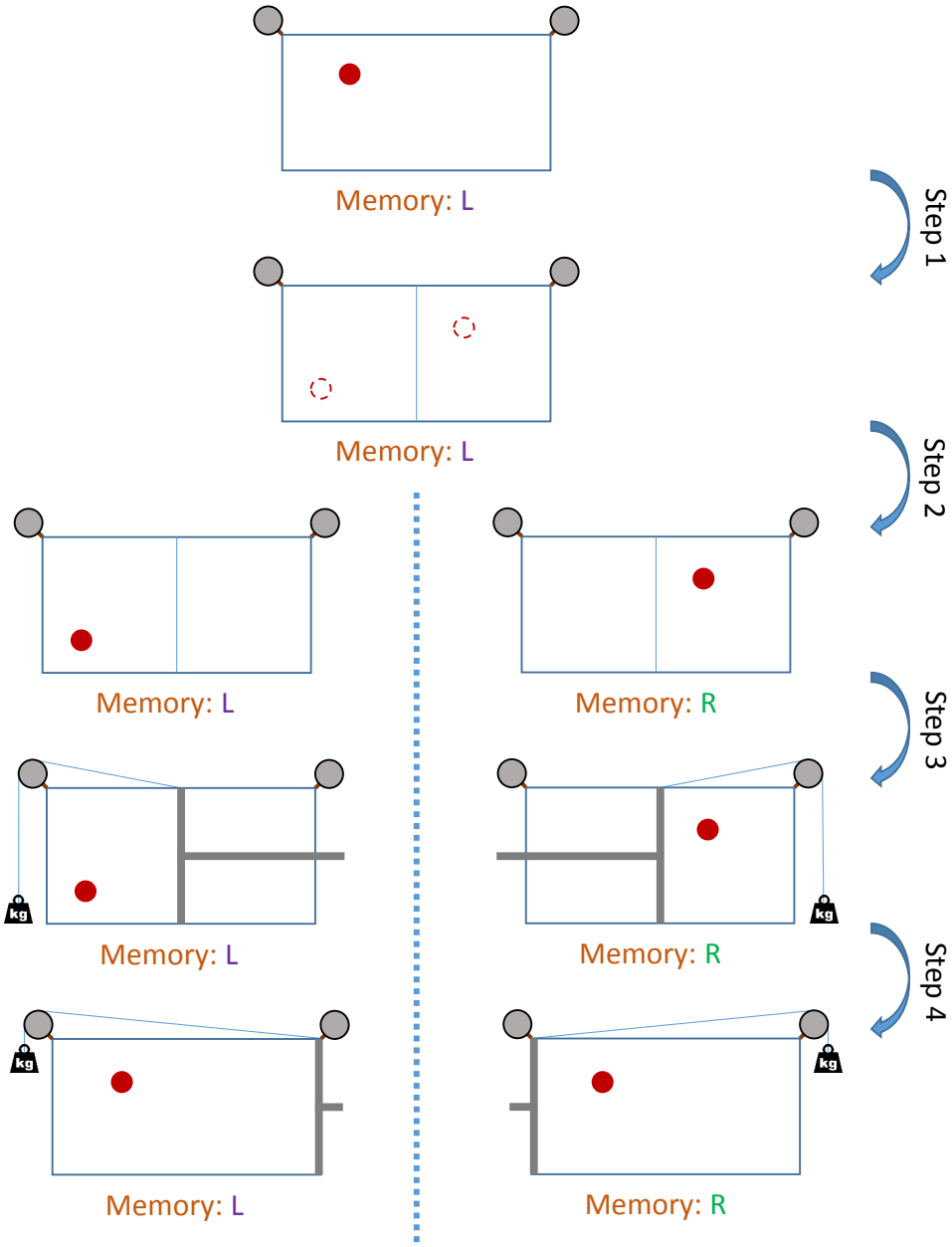



Figure 4.1: *The Szilard engine.* Here we show the 4 steps for work extraction from a Szilard engine as described in the main text. Note that while the system returns to its original state, the memory does not.

Within a resource theory, one is often concerned with the task of converting one state into another and this is usually investigated in two regimes: when one has access to a large number of copies of an initial resource state and when one has access to only a single copy. In the first scenario, relevant questions pertain to the *rate* at which one state can be converted into another. Given $n$ copies of $\rho$, where $n$ is a large number, what fraction of them can be converted into copies of some other $\sigma$ using $\mathcal{C}$? Such transformations can allow one to trade copies of one resource state for fewer copies of another, more useful one.

It is the second scenario that we will focus on in this thesis as we are interested in thermodynamics when one does not have access to many copies of a particular state. Here one asks: given a single copy of $\rho$, is it possible to produce a single copy of $\sigma$ using $\mathcal{C}$? Such a regime is often referred to as *single-shot* within information theory but for our purposes it equates to not taking the thermodynamic limit. If $\sigma$ belongs to $\mathcal{S}$, then the answer will always be yes. For instances where $\sigma$ cannot be produced from $\rho$, then it may be possible to drive the transition by supplying some additional amount of resource. Conversely, if $\sigma$ can be produced from $\rho$, can one in addition create another resource state? In both cases, how should these amounts of resource be quantified and optimized? In answering the questions in this paragraph, it is often useful to determine the *monotones* of the resource theory. These are functions of states that never increase under the action of $\mathcal{C}$. Hence, if the value of a monotone evaluated on the target state is larger than its value evaluated on the initial state, the transition is not possible under $\mathcal{C}$.

Perhaps the most well studied example of a quantum resource theory, is that of bipartite pure state manipulation under local operations and classical communication (LOCC). Within this paradigm, product states come for free as, using LOCC, they can always be created. On the other hand, entangled states cannot be created from product states under this class of operations and hence, can be regarded as resources. Indeed, they can be used to implement operations such as teleportation which are beyond the reach of LOCC. In general, there exist necessary and sufficient conditions for a given single copy (single-shot) transformation to be possible under LOCC [133]. These conditions take the form of *majorization* relations - a concept that will be of vital importance in our consideration of single-shot thermodynamics.

The resource theory approach has also been applied to topics as diverse as asymmetry [78], steering [69] and coherence manipulation [170], and their general structure has been studied [23]. With respect to nano-scale thermodynamics, two resource theories will be of interest.

The first, noisy operations [88], is a resource theory for purity manipulation and can also be applied to study the thermodynamics of systems with trivial Hamiltonians. As such it often provides insight for dealing with the non-trivial case which is the subject of the resource theory of thermal operations [96, 89].

### 4.2.1 Noisy operations

In the resource theory of noisy operations (NO) [88], three actions can be applied to a system in a given state, $\rho$:

1. A system of any dimension, in the maximally mixed state can be appended.
2. Any unitary can be applied to the global system.
3. Any subsystem can be discarded through tracing out.

Combining these, the action of a noisy operation on $\rho$ can be written as:

$$\rho \xrightarrow{\text{NO}} \text{Tr}_{A'} \left[ U \left( \rho \otimes \frac{\mathbb{I}_A}{d_A} \right) U^\dagger \right], \tag{4.4}$$

where $A$ is an ancillary system of dimension $d_A$ appended in the maximally mixed state, $U$ is a unitary acting on both the system and the ancilla and $A'$ is the subsystem to be discarded. Within this theory, maximally mixed states are the free states as, by using Operation 1, they can always be created. Any state with a degree of purity (that is, it is not a maximally mixed state) is a resource. A comprehensive study of this resource theory can be found in [77] which covers the monotones of the theory and both exact and inexact state conversion including in the presence of a catalyst (a topic we shall discuss briefly in Section 4.3.1).

#### Majorization

Given access to these three operations, when is it possible to transform a system in state $\rho$ into a system in state $\sigma$? The answer to the question can be stated in terms of *majorization*:

**Definition 23** (Majorization)**.** *Given two normalized probability distributions, $p$ and $q$, each consisting of $n$ elements, let $\vec{p} = \{p_1, \ldots, p_n\}$ and $\vec{q} = \{q_1, \ldots, q_n\}$ represent their elements written in non-increasing order. We say $p$ majorizes $q$, written $p \succ q$, if:*

$$\sum_{i=1}^{k} p_i \geq \sum_{i=1}^{k} q_i, \quad \forall k \in \{1, \ldots, n\}. \tag{4.5}$$

*For two quantum states, $\rho$ and $\sigma$, on a Hilbert space of dimension $n$, let $\vec{\eta} = \{\eta_1, \ldots, \eta_n\}$ denote the eigenvalues of $\rho$, listed in non-increasing order, and $\vec{\zeta} = \{\zeta_1, \ldots, \zeta_n\}$ denote the eigenvalues of $\sigma$, similarly ordered. We say $\rho$ majorizes $\sigma$, written $\rho \succ \sigma$, if $\vec{\eta} \succ \vec{\zeta}$.*

The canonical textbook on the subject of majorization is [120]. In order to answer the question posed at the start of the previous paragraph we will require two more definitions and lemmas:

**Definition 24** (Bistochastic, completely positive, linear map)**.** *A completely positive linear map is called bistochastic if it is trace preserving and maps the identity to itself.*[1]

The relevance of such maps to majorization is given by the following lemma:

**Lemma 11.** *[41]. Given two quantum states $\rho$ and $\sigma$ on a Hilbert space of dimension $n$, $\rho$ majorizes $\sigma$ if and only if there exists a bistochastic, completely positive, linear map that transforms $\rho$ into $\sigma$.*

In addition we shall define a particularly simple type of transformation that acts on only two components of a probability distribution:

**Definition 25** (T-transforms)**.** *A T-transform is a linear map whose matrix, $T$, can be written in the form:*

$$T = \lambda \mathbb{I} + (1 - \lambda)\, Q, \tag{4.6}$$

*where $0 \leq \lambda \leq 1$ and $Q$ is permutation matrix that interchanges two coordinates.*

*The action of $T$ on the vector $p = (p_1, \ldots, p_n)$ is given by:*

$$Tp = (p_1, \ldots, \lambda p_i + (1 - \lambda)\, p_j, \ldots, \lambda p_j + (1 - \lambda)\, p_i, \ldots, p_n)\,, \tag{4.7}$$

*where $i$ and $j$ denote the coordinates permuted by $Q$.*

These relate to majorization through:

**Lemma 12.** *[128, 80]. Given two probability distributions, $p$ and $q$, if $p$ majorizes $q$, then $p$ can be converted into $q$ by successive applications of a finite number of T-transforms.*

With these in place, we can now derive a necessary and sufficient condition from [88] for it to be possible to transform $\rho$ into $\sigma$ under noisy operations.

[1] Note that such maps act via bistochastic matrices on the eigenvalues of the quantum state.

**Theorem 15.** *[88]. Given two states, $\rho$ and $\sigma$, of an n-level system,*[2] *then:*

$$\rho \xrightarrow{NO} \sigma, \tag{4.8}$$

*if and only if $\rho \succ \sigma$.*

*Proof.* That $\rho \xrightarrow{\text{NO}} \sigma$ implies $\rho \succ \sigma$, follows readily from Lemma 11 and the fact that all noisy operations are bistochastic, completely positive, linear maps.

To show the reverse implication, we deviate from the proof given in [88]. First note that as we can apply any unitary, we can always assume that $\rho$ and $\sigma$ commute. Our goal is to show that there exists a noisy operation that implements an arbitrary T-transform on the eigenvalues of $\rho$. Then, from Lemma 12, the result will follow. We wish to show that given a state, $\gamma$, of a 2-level system, it is possible to transform:

$$\gamma = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix} \xrightarrow{\text{NO}} \begin{pmatrix} \lambda\gamma_1 + (1-\lambda)\,\gamma_2 & 0 \\ 0 & \lambda\gamma_2 + (1-\lambda)\,\gamma_1, \end{pmatrix}$$

for any $0 \leq \lambda \leq 1$.

To do this, assume that $\lambda = \frac{j}{d}$ (if $\lambda$ is irrational, then $j$ and $d$ should be chosen such that the T-transform is implemented to the desired accuracy). A noisy operation protocol to implement the T-transform is then as follows:

- Step 1. Append an ancilla of dimension $d$ in the maximally mixed state:

$$\gamma = \text{diag}\,(\gamma_1, \gamma_2) \longrightarrow \frac{1}{d}\text{diag}\big(\underbrace{\gamma_1, \ldots \gamma_1}_{d}, \underbrace{\gamma_2, \ldots, \gamma_2}_{d}\big).$$

- Step 2. Perform a unitary that swaps $d-j$ of the $\frac{\gamma_2}{d}$ with $\frac{\gamma_1}{d}$:

$$\frac{1}{d}\text{diag}\,(\gamma_1, \ldots \gamma_1, \gamma_2, \ldots, \gamma_2) \longrightarrow \frac{1}{d}\text{diag}\big(\underbrace{\gamma_1, \ldots, \gamma_1}_{j}, \underbrace{\gamma_2, \ldots, \gamma_2}_{d-j}, \underbrace{\gamma_2, \ldots, \gamma_2}_{j}, \underbrace{\gamma_1, \ldots, \gamma_1}_{d-j}\big).$$

- Step 3. Discard the ancilla system:

$$\begin{aligned} &\frac{1}{d}\text{diag}\big(\underbrace{\gamma_1, \ldots, \gamma_1}_{j}, \underbrace{\gamma_2, \ldots, \gamma_2}_{d-j}, \underbrace{\gamma_2, \ldots, \gamma_2}_{j}, \underbrace{\gamma_1, \ldots, \gamma_1}_{d-j}\big) \\ &\qquad \longrightarrow \text{diag}\left(\frac{j}{d}\gamma_1 + \frac{d-j}{d}\gamma_2, \frac{j}{d}\gamma_2 + \frac{d-j}{d}\gamma_1\right), \end{aligned}$$

  as required.

[2]Note that we can always assume that $\rho$ and $\sigma$ have the same dimension. If they do not, by applying Operation 1 of noisy operations appropriately, we can ensure that the systems under consideration have the same dimension.

Hence, $\rho \succ \sigma$ implies that $\rho \xrightarrow{\text{NO}} \sigma$. □

A valuable tool for visualizing the criteria provided in Theorem 15 is that of Lorenz curves, illustrated in Figure 4.2. We will regularly use these (and their equivalent constructions in the context of thermal operations) throughout this thesis to provide intuition and to argue about which transformations are allowed within the resource theory.

**Definition 26** (Lorenz curves)**.** *For a given state, $\rho$, with ordered eigenvalues $\vec{\eta}$, we define the functions $V_k$, for $k \in \{1, \ldots, n\}$, by:*

$$V_k(\rho) = \sum_{i=1}^{k} \eta_i. \tag{4.9}$$

*The Lorenz curve for $\rho$ is then formed by plotting the points:*

$$\left\{ \left( \frac{k}{n}, V_k(\rho) \right) \right\}_{k=1}^{n}, \tag{4.10}$$

*together with the point $(0, 0)$, and connecting them piecewise linearly to form a concave curve.*

*We call a point from the set given in Eq.* (4.10) *an elbow, if the gradient of the Lorenz curve changes as it passes through that point. Otherwise, it is a non-elbow.*

If $\rho$ majorizes $\sigma$, the Lorenz curve of $\rho$ is never below that of $\sigma$ and:

$$V_l(\rho) \geq V_l(\sigma), \quad \forall l \in \{1, \ldots, n\}. \tag{4.11}$$

The functions defined in Eq. (4.9), together with their analogue in thermal operations, will be crucial for deriving the results of Chapter 5. They are monotones of the theory, only decreasing under noisy operations.

### Sharp states

If it is not possible to deterministically convert $\rho$ into $\sigma$ using noisy operations, to perform the transformation with certainty will require the consumption of an additional resource state containing purity. Likewise, if $\rho$ can be converted into $\sigma$ deterministically, it may be possible to extract some resource state from the process. The amount of purity that is required or can be produced can be quantified using *sharp states* [77]. A sharp state of a $d$-level system has rank

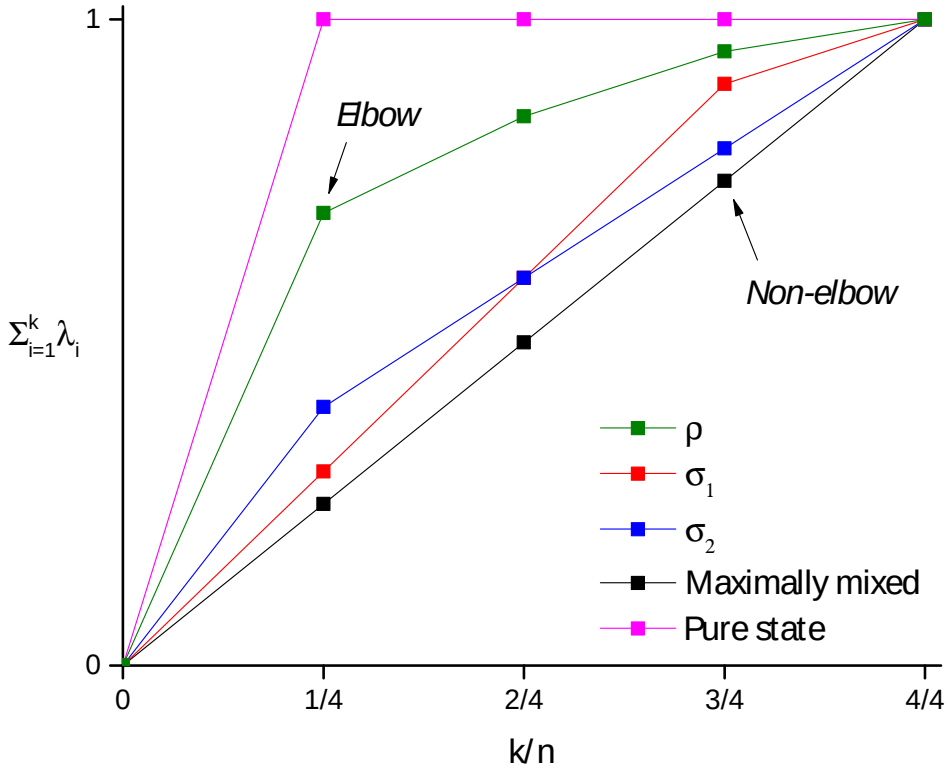



Figure 4.2: *Lorenz curves.* *a)* The Lorenz curve for $\rho$ is defined by plotting the points: $\left\{\left(\frac{k}{n}, V_k\left(\rho\right)\right)\right\}_{k=1}^{n}$. *b)* It is possible to transform $\rho$ into either of $\sigma_1$ or $\sigma_2$ using noisy operations as the Lorenz curve of $\rho$ is never below the Lorenz curves of $\sigma_1$ and $\sigma_2$. *c)* It is not possible to transform $\sigma_1$ into $\sigma_2$ or $\sigma_2$ into $\sigma_1$ using noisy operations as their Lorenz curves cross. *d)* The Lorenz curve of a maximally mixed state is given by a straight line between $(0, 0)$ and $(1, 1)$. All other states majorize it. *e)* A pure state of an $n$-level system majorizes all other $n$-level system states and corresponds to the sharp state $s_{\log_2 n}$ (see Eq. (4.12)). *f)* We call a point on a Lorenz curve an elbow, if the gradient of the curve changes as it passes through the point. Otherwise, it is a non-elbow.

$j$ and is maximally mixed on its support. We denote such a state by $s_{\log_2 \frac{d}{j}}$, and its density matrix is:

$$s_{\log_2 \frac{d}{j}} = \operatorname{diag}\Big(\underbrace{\frac{1}{j}, \ldots, \frac{1}{j}}_{j}, \underbrace{0, \ldots, 0}_{d-j}\Big). \tag{4.12}$$

With this in place, we denote the amount of purity consumed/produced in the transition by $S_{\rho\to\sigma}$ and define it to be the greatest value of $S$ such that one of the following holds:

$$\begin{aligned} \rho \otimes s_{|S|} &\xrightarrow{NO} \sigma, && \text{if } S \leq 0, \\ \rho &\xrightarrow{NO} \sigma \otimes s_{|S|}, && \text{if } S > 0. \end{aligned} \tag{4.13}$$

If $S_{\rho\to\sigma}$ is negative, additional purity has been consumed in converting $\rho$ into $\sigma$ while if it is positive, some extra purity has been produced. In terms of Lorenz curves, tensoring a state $\rho$ with a sharp state $s_z$, has the effect of compressing the Lorenz curve of $\rho$ by a factor of $2^{-z}$ with respect to the $x$-axis.

We shall investigate $S_{\rho\to\sigma}$ further in Chapter 5. For now we define the *distillable purity* and the *purity of formation*. The first of these quantities is the maximum amount of purity we can extract in transforming $\rho$ into a maximally mixed state and given by:

$$S_{\text{distil}}(\rho) = \log_2 \left( \frac{n}{\operatorname{rank}(\rho)} \right). \tag{4.14}$$

It is always non-negative. Similarly, the purity of formation is defined as the minimum amount of purity required to create $\rho$ if we started from a maximally mixed state. It is never positive and is given by:

$$S_{\text{form}}(\rho) = -\log_2(\eta_1 n), \tag{4.15}$$

where $n$ is the dimension of $\rho$ and $\eta_1$ its largest eigenvalue.

The sharp states related to $S_{\text{distil}}(\rho)$ and $S_{\text{form}}(\rho)$ are shown in Figure 4.3, together with their relation to the Lorenz curve of $\rho$. Note that in general, $|S_{\text{distil}}(\rho)| \leq |S_{\text{form}}(\rho)|$ with equality if and only if $\rho$ is a sharp state. This means that noisy operations is not a reversible theory at the level of single-shot transitions as the purity required to form $\rho$ is greater than the purity that can be distilled from it.

### 4.2.2 Thermal operations

Noisy operations can be generalized to create a resource theory for the thermodynamics of systems with arbitrary, finite Hamiltonians. This leads to the resource theory of thermal operations (TO) [96, 89], which models the thermodynamics of a system in the presence of a

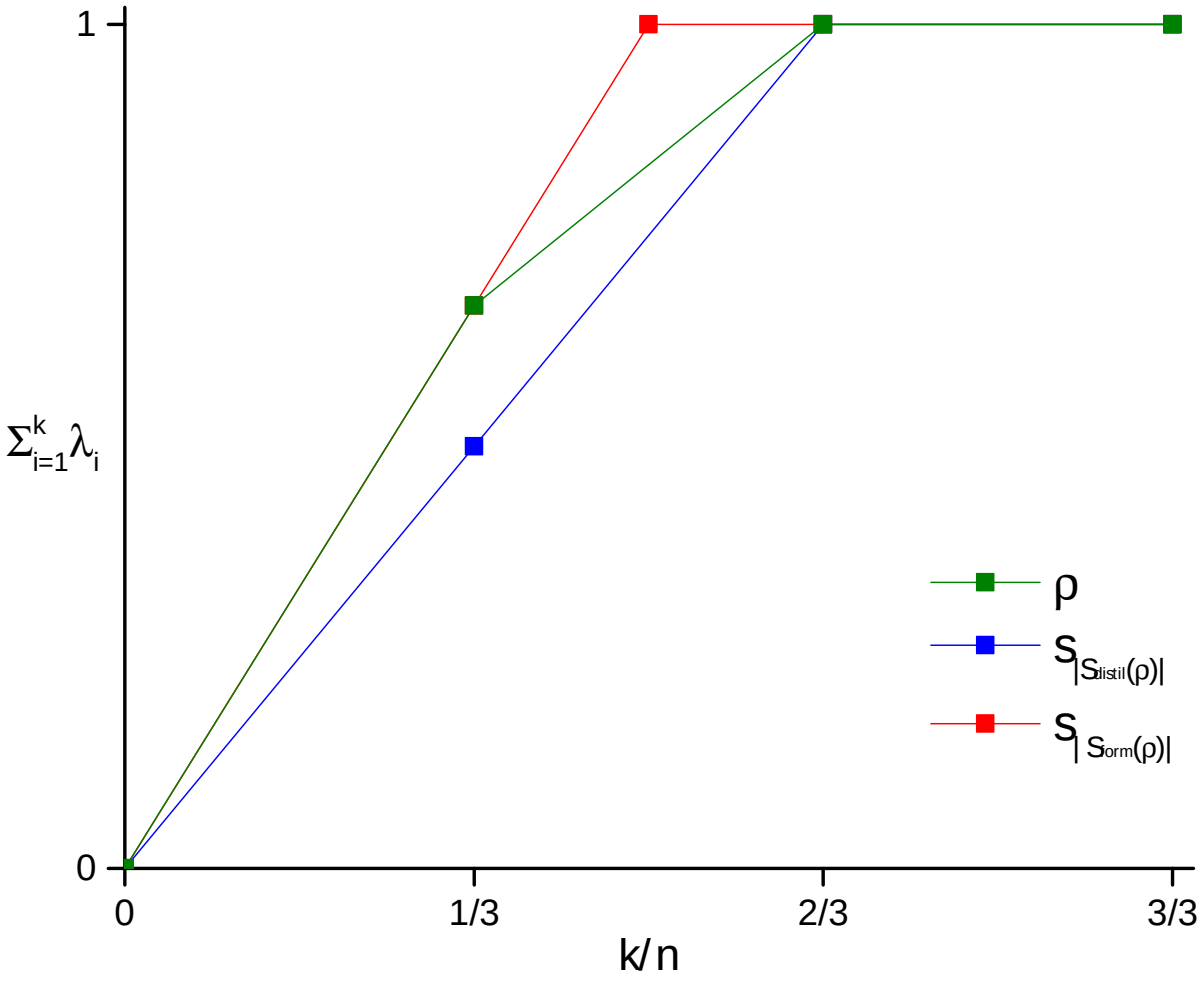



Figure 4.3: *Purity of formation and distillation.* $S_{\text{distil}}(\rho)$ is characterized by the sharpest state majorized by $\rho$ and $S_{\text{form}}(\rho)$ is characterized by the least sharp state that majorizes $\rho$.

large heat bath. Given a system in state $\rho$, with Hamiltonian $H_S$ and such a heat bath at temperature $T$, the following operations can be applied:

1. A system with any Hamiltonian, in the Gibbs state of that Hamiltonian at temperature $T$, can be appended as an ancilla. For the Hamiltonian $H_B = \sum_{i=1}^{n} E_i |i\rangle\langle i|$, the corresponding Gibbs state at temperature $T$ is:

$$\tau_B = \frac{1}{Z_B} \sum_{i=1}^{n} e^{-\beta E_i} |i\rangle\langle i|, \tag{4.16}$$

   where $\beta = \frac{1}{k_B T}$ is the inverse temperature and $Z_B = \sum_{i=1}^{n} e^{-\beta E_i}$ is the partition function.

2. Any energy conserving unitary, i.e. those unitaries that commute with the total Hamiltonian, can be applied to the global system.

3. Any subsystem can be discarded through tracing out.

Given these, the action of a thermal operation on the state, $\rho$, of a state-Hamiltonian pair $(\rho, H_S)$ at temperature $T$, can be written as:

$$\rho \xrightarrow{\text{TO}} \text{Tr}_A \left[ U \left( \rho \otimes \tau_B \right) U^\dagger \right], \tag{4.17}$$

where $\tau_B$ is the Gibbs state of a Hamiltonian $H_B$, $A$ is the subsystem to be discarded and $U$ is a unitary such that:

$$[U, H_S + H_B] = 0, \tag{4.18}$$

and we have used the shorthand notation $H_S + H_B = H_S \otimes \mathbb{I}_B + \mathbb{I}_S \otimes H_B$.

The Gibbs state of $H_S$ at temperature $T$ is the free state of the theory. It can always be created as an ancilla by using Operation 1 and then swapped with the system using an energy conserving unitary before discarding the ancillary system. Indeed, Gibbs states are the only choice of state that can be appended for free under Operation 1 without making any transition between states possible and the entire theory becoming trivial [24].

By demanding that all interactions between the system and bath are governed by energy conserving unitaries, modeling thermodynamics using thermal operations enables us to precisely keep track of the energy, entropy and work manipulated during a given process. Requiring that the allowed unitaries obey Eq. (4.18) also ensures that energy is conserved on every possible state given as input to a thermal operation [89].

In spite of these constraints, the operations allowed under this regime are still very general and not necessarily easy to implement experimentally (though we will discuss this further in Chapter 6). As such, bounds or constraints derived using them can be regarded as truly fundamental. Note also, that a number of other paradigms that one can use to study thermodynamics, for example allowing interaction Hamiltonians or changing energy levels, can be modeled using the thermal operations framework [89, 25].

Using Operation 1 of thermal operations, a particularly useful, ideal, heat bath can be created [89]. Properties of this ideal heat bath such as the energy and degeneracies are taken to be large, tending to infinite, while in comparison, properties of the system we are trying to manipulate are relatively small. Labeling the bath system by $R$, it is taken to be in a Gibbs state, $\tau_R$, at temperature $T$ that with high probability occupies energies from a set, $\mathcal{E}_R$. For energies within this set, the following properties hold:

1. The energies $E \in \mathcal{E}_R$ are peaked around a mean value. They satisfy:

$$E \in \left\{ \langle E \rangle - O\left(\sqrt{\langle E \rangle}\right), \ldots, \langle E \rangle + O\left(\sqrt{\langle E \rangle}\right) \right\}. \tag{4.19}$$

   This is a generic property of most heat baths.

2. For $E \in \mathcal{E}_R$, the degeneracies, $g_R(E)$, satisfy:

$$g_R(E) \geq e^{cE}, \tag{4.20}$$

   for some constant $c$.

3. The energy spectrum of the bath is assumed to be such that for all $E \in \mathcal{E}_R$ and distinct energy levels of the system, $E_i$ and $E_j$, there exists $E' \in \mathcal{E}_R$ such that $E + E_i = E' + E_j$.

4. For $E \in \mathcal{E}_R$, the degeneracies are such that:

$$g_R(E - E_i) \approx g_R(E)\, e^{-\beta E_i}, \tag{4.21}$$

   where $E_i$ is an energy level of the system under consideration.

Note that such a heat bath can be obtained by using Operation 1 of thermal operations to create the state $\tau^{\otimes n}$, for large $n$. Furthermore, if one has access to a heat bath with these properties, then adding a thermal state of relatively small dimension to it using this operation produces a larger heat bath that still satisfies these assumptions [89].

With thermal operations defined and these properties of an ideal heat bath in place, we now turn to *thermo-majorization*, a concept akin to that of majorization used in noisy operations.

### Thermo-majorization

When is it possible to transform the state of a system into another under thermal operations? We begin by considering processes in which the Hamiltonian of the system does not change. As shown in [89], determining whether a transition form $(\rho, H_S)$ to $(\sigma, H_S)$ is achievable can be formulated using *thermo-majorization diagrams*. These are similar to the Lorenz curves of noisy operations but with two crucial differences. Suppose $\rho$ is block-diagonal in the energy eigenbasis (i.e. it contains no coherence between energy eigenspaces) with eigenvalue $\eta_i$ associated with each energy level $E_i$ of an $n$-level system. Then firstly, rather than ordering the eigenvalues according to the magnitude of $\eta_i$, we instead *$\beta$-order* them, listing them such that $\eta_i e^{\beta E_i}$ is in non-increasing order.

The second difference is that we no longer plot the $\beta$-ordered $\eta_i$ at evenly spaced intervals. Instead, we plot the points:

$$\left\{ \sum_{i=1}^{k} e^{-\beta E_i^{(\rho)}}, \sum_{i=1}^{k} \eta_i^{(\rho)} \right\}_{k=1}^{n}, \tag{4.22}$$

where the superscript $\rho$ on $E_i$ and $\eta_i$ indicated that they have been $\beta$-ordered and this ordering depends on $\rho$. We say that one block-diagonal state *thermo-majorizes* another if its thermo-majorization curve never lies below that of the other. This is illustrated in Figure 4.4.

If a state, $\rho$, is not block-diagonal in the energy eigenbasis and has coherences between energy levels, determining whether a transition is possible or not is written in terms of the state formed by decohering $\rho$ in the energy eigenbasis. This state, $\rho_D$, is given by:

$$\rho_D = \sum_{i=1}^{n} |i\rangle\langle i|\rho|i\rangle\langle i|, \tag{4.23}$$

where $\{|i\rangle\}_{i=1}^{n}$ are a set of orthonormal eigenvectors of the system's Hamiltonian. The operation of decohering $\rho$ to give $\rho_D$ can be performed using a thermal operation that commutes with all other thermal operations [25]. We shall define the thermo-majorization curve of a state with coherences to be the thermo-majorization curve of that state decohered in the energy eigenbasis as per Eq. (4.23).

With these definitions in place, we can give the result of [89]:

**Theorem 16.** *[89]. Given two states, $\rho$ and $\sigma$, of an $n$-level system with Hamiltonian $H_S$:*

1. *If $\rho$ and $\sigma$ are block-diagonal in the energy eigenbasis, then $\rho \xrightarrow{TO} \sigma$, if and only if $\rho$ thermo-majorizes $\sigma$.*

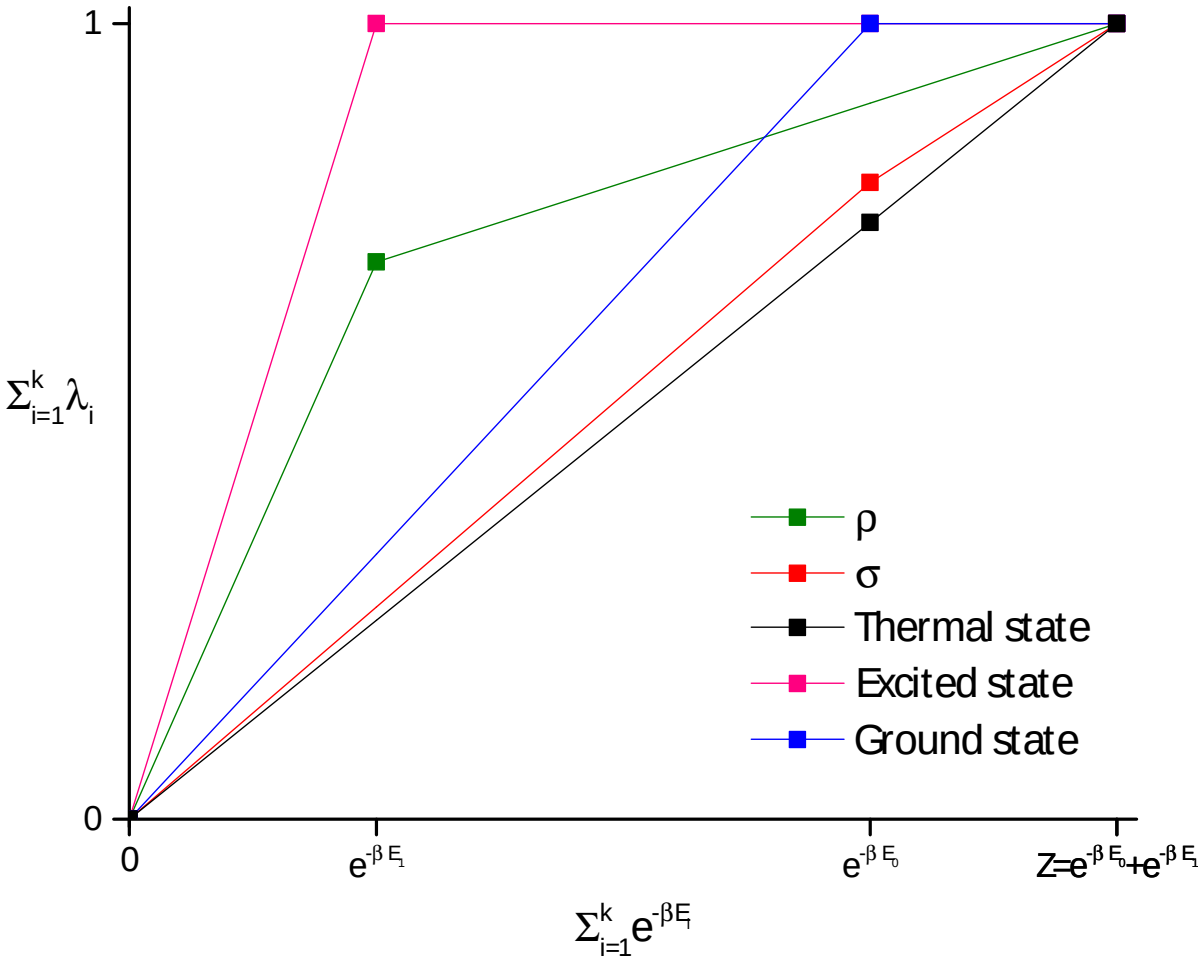



Figure 4.4: *Thermo-majorization curves.* *a)* The thermo-majorization curve for a block-diagonal state, $\rho$, is defined by plotting the points: $\left\{\sum_{i=1}^{k} e^{-\beta E_i^{(\rho)}}, \sum_{i=1}^{k} \eta_i^{(\rho)}\right\}_{k=1}^{n}$. Note that states associated with the same Hamiltonian may have different $\beta$-orderings, as illustrated by $\rho$ and $\sigma$ here. *b)* Here, $\rho$ thermo-majorizes $\sigma$ as the thermo-majorization curve of $\rho$ is never below the thermo-majorization curve of $\sigma$. *c)* The thermo-majorization curve of a Gibbs state is given by a straight line between $(0,0)$ and $(Z,1)$. All other states thermo-majorize it. *d)* The pure state corresponding to the highest energy level of an $n$-level system thermo-majorizes all other states associated with that Hamiltonian.

2. *If $\sigma$ is block-diagonal in the energy eigenbasis, then $\rho \xrightarrow{TO} \sigma$, if and only if $\rho_D$ thermo-majorizes $\sigma$.*

3. *In general, $\rho \xrightarrow{TO} \sigma$, only if $\rho_D$ thermo-majorizes $\sigma_D$.*

*Proof.* We shall only sketch the proof here and more precise statements can be found in [89]. The main idea is to transform the problem into the framework of noisy operations where Theorem 15 applies. For simplicity, we shall consider the case where $\rho$ and $\sigma$ are block-diagonal in the energy eigenbasis.

Let $\rho = \sum_{i=1}^{n} \eta_i |i\rangle\langle i|$, $\sigma = \sum_{i=1}^{n} \zeta_i |i\rangle\langle i|$ and $H_S = \sum_{i=1}^{n} E_i |i\rangle\langle i|$. Let $\tau_R$ be the state of the large heat bath, possessing the four properties given in Section 4.2.2. As noted previously, this can be created under thermal operations. Consider fixing the total energy of the system and bath to be $E$. If $P_E$ denotes the projection of the joint system onto states with total energy $E$, then using heat bath Properties 1-3, it can be justified [89] that $P_E \left(\rho \otimes \tau_R\right) P_E$ is close to a state with eigenvalues:

$$e^{\beta E_i} \frac{\eta_i}{g_R\left(E\right)} \text{ with degeneracy } e^{-\beta E_i} g_R\left(E\right),$$

where $g_R\left(E\right)$ denotes the degeneracy of the bath energy levels and we have made use of heat bath Property 4. The state $P_E \left(\sigma \otimes \tau_R\right) P_E$ has eigenvalues and degeneracies similarly defined.

On joint states with total energy $E$, the action of thermal operations is similar to that of noisy operations. Hence, $P_E \left(\rho \otimes \tau_R\right) P_E$ can be converted into $P_E \left(\sigma \otimes \tau_R\right) P_E$ if and only if the majorization relation of Theorem 15 holds. We thus arrange the eigenvalues of $P_E \left(\rho \otimes \tau_R\right) P_E$ such that $\left\{e^{\beta E_i} \eta_i\right\}_{i=1}^{n}$ is in descending order (essentially $\beta$-ordering the $\left\{\eta_i\right\}_{i=1}^{n}$) and consider the associated Lorenz curve. This is illustrated in Figure 4.5.

As the majorization order obtained from such Lorenz curves does not depend on the total energy $E$, we obtain that $\rho \xrightarrow{\text{TO}} \sigma$ if and only if $\rho$ thermo-majorizes $\sigma$. □

Similarly to how Eq. (4.9) defines monotones for the resource theory of noisy operations, the height of a thermo-majorization curve provides monotones for thermal operations. If we denote the height of the thermo-majorization curve of $\rho$ at $x$ by $\tilde{V}_x\left(\rho\right)$, for $0 \leq x \leq Z$, then by Theorem 16, these functions are non-increasing under thermal operations. In particular, for block-diagonal $\rho$, we have:

$$\tilde{V}_{x_k}\left(\rho\right) = \sum_{i=1}^{k} \eta_i^{(\rho)}, \quad \text{where } x_k = \sum_{i=1}^{k} e^{-\beta E_i^{(\rho)}}. \tag{4.24}$$

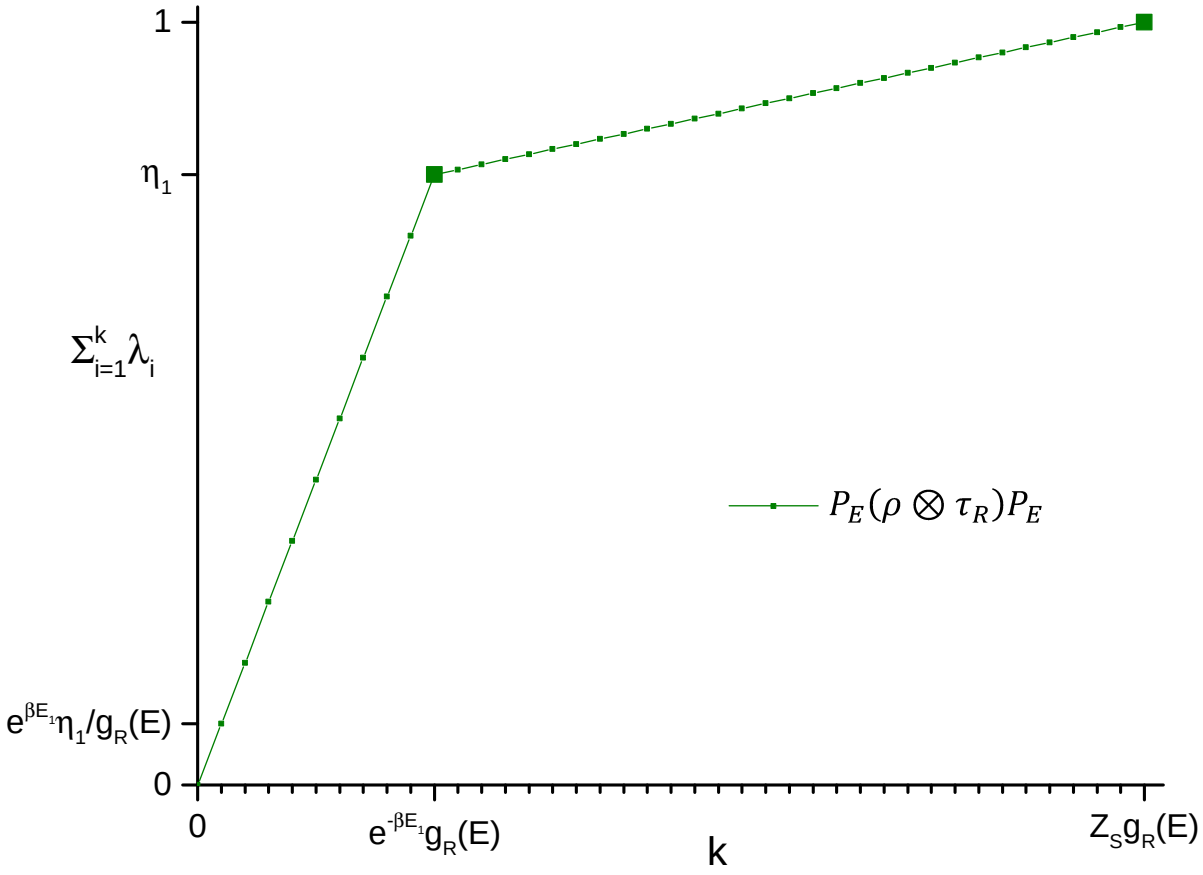



Figure 4.5: *Lorenz curve for* $P_E\left(\rho \otimes \tau_R\right) P_E$. Here we illustrate how thermo-majorization curves arise from considering the Lorenz curve of $\rho \otimes \tau_R$ restricted to fixed energy, $E$. Note that the structure of the curve (the relative position of the elbows) does not depend on $E$.

These monotones also give us an alternative way of stating the thermo-majorization criteria:

**Theorem 17.** *Suppose $\sigma$ is block-diagonal in the energy eigenbasis and let:*

$$\mathcal{L}(\sigma) = \left\{ \sum_{i=1}^{k} e^{-\beta E_i^{(\sigma)}} \right\}_{k=1}^{n} . \tag{4.25}$$

*Then a state $\rho$ can be deterministically converted into $\sigma$ under thermal operations if and only if:*

$$\tilde{V}_x(\rho_D) \geq \tilde{V}_x(\sigma), \quad \forall x \in \mathcal{L}(\sigma) . \tag{4.26}$$

*Proof.* Suppose $\rho \xrightarrow{TO} \sigma$. Then by Theorem 16, $\tilde{V}_x(\rho_D) \geq \tilde{V}_x(\sigma)$, for $0 \leq x \leq Z$ and in particular, Eq. (4.26) holds.

Conversely, suppose Eq. (4.26) holds and, setting $t_0 = 0$, label the elements of $\mathcal{L}(\sigma)$ arranged in increasing order by $t_i$, for $i = 1$ to $n$. Then on the interval $[t_{i-1}, t_i]$, for $i \in \{1, \ldots, n\}$, the thermo-majorization curve of $\sigma$ is given by a straight line. From $\rho$, define the block-diagonal state $\rho_\sigma$ by the thermo-majorization curve:

$$\left\{ \left( t_i, \tilde{V}_{t_i}(\rho_D) \right) \right\}_{i=1}^{n} , \tag{4.27}$$

and note that due to the concavity of thermo-majorization curves, $\rho_D$ thermo-majorizes $\rho_\sigma$. On the interval $[t_{i-1}, t_i]$, $i \in \{1, \ldots, n\}$, the thermo-majorization curve of $\rho_\sigma$ is also given by a straight line. The construction of $\rho_\sigma$ is shown in Figure 4.6.

As by construction $\tilde{V}_{t_i}(\rho_\sigma) = \tilde{V}_{t_i}(\rho_D)$, for all $i$, Eq. (4.26) implies that $\tilde{V}_{t_i}(\rho_\sigma) \geq \tilde{V}_{t_i}(\sigma)$, for all $i$. Hence on each interval $[t_{i-1}, t_i]$, the thermo-majorization curves for $\rho_\sigma$ and $\sigma$, and therefore $\rho$ and $\sigma$, do not cross. As this holds for all $i$ and the intervals cover $[0, Z]$, the thermo-majorization curve of $\rho$ is never below that of $\sigma$ and we can perform $\rho \xrightarrow{TO} \sigma$ deterministically. □

If the number of 'elbows' in the thermo-majorization curve of $\sigma$ is $j$, then this theorem reduces thermo-majorization to checking $j$ criteria and generalizes Lemma 17 of [77] (an analogous statement for noisy operations) to thermal operations. Note also that if $\sigma$ is not block-diagonal in the energy eigenbasis, replacing $\sigma$ by $\sigma_D$ in Eqs. (4.25) and (4.26) gives a necessary but not sufficient condition for the transition from $\rho$ to $\sigma$ to be possible. This is as strong a constraint as that given in Theorem 16 Part 3.

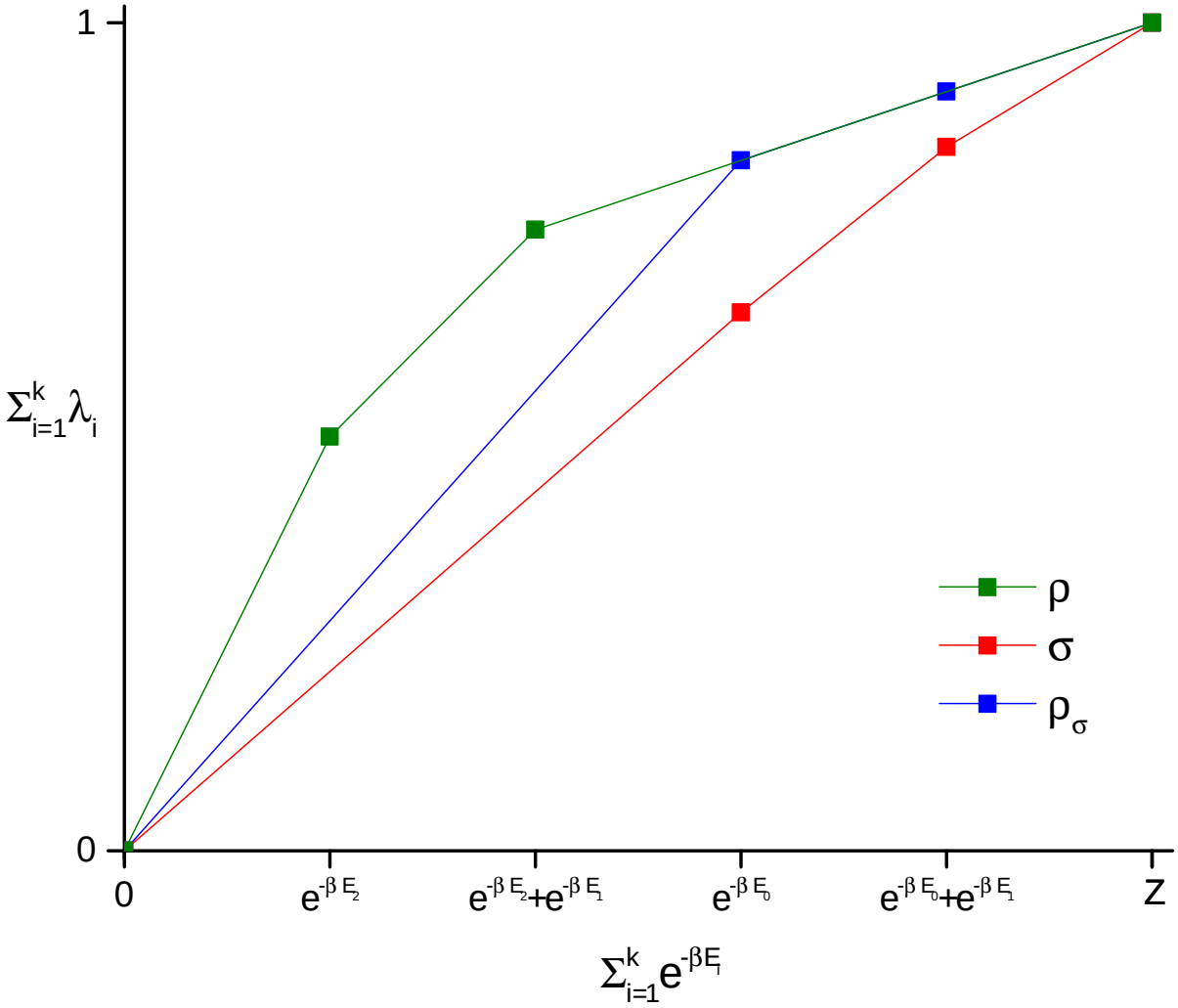



Figure 4.6: *The construction of* $\rho_\sigma$. To form the thermo-majorization curve of $\rho_\sigma$, the points on the curve of $\rho$ that are at the same horizontal position as the elbows of $\sigma$ are joined. By concavity, the resultant curve is always below the thermo-majorization curve of $\rho$.

### The work of transition

In general, if we want a transition from $\rho$ to $\sigma$ to be possible, an additional resource, *work*, may have to be supplied. Alternatively, if a transition can be achieved with certainty, it can be possible to extract work. In traditional thermodynamics, work is often represented and measured by considering a work storage system consisting of a suspended weight. During a given process, if the weight is raised, work has been stored in system for use at a later date while if it is lowered, work has been expended in performing the procedure.

Within the framework of thermal operations, the optimal amount of work that must be added or gained, the *work of transition*, can be quantified using the energy gap, $W$, of a 2-level system with zero-energy state, $|0\rangle$, and an additional state, $|1\rangle$. The associated Hamiltonian is then:

$$H_W = W|1\rangle\langle 1|. \tag{4.28}$$

The work of transition, denoted $W_{\rho\to\sigma}$, is then defined to be the greatest value of $W$ such that:

$$(\rho \otimes |0\rangle\langle 0|, H_S + H_W) \xrightarrow{\text{TO}} (\sigma \otimes |1\rangle\langle 1|, H_S + H_W)\,. \tag{4.29}$$

If $W_{\rho\to\sigma}$ is negative, to convert $\rho$ into $\sigma$ work has been taken from the work system to enable the transition to take place. On the other hand, if $W_{\rho\to\sigma}$ is positive, then in converting $\rho$ into $\sigma$ it has been possible store some extracted work.

Defining work in such a way enables the quantification of the *worst-case work* of a process. When $W_{\rho\to\sigma}$ is negative, it can be interpreted as the smallest amount of work that must be supplied to ensure the transition takes place. If it is positive, it is the largest amount of work we are guaranteed to extract in the process. As the work system is both initially and finally in a pure state, no entropy is contained within it and its energy change must be completely due to work being exchanged with the system.

Figure 4.7 shows the effect of appending the work storage system in either of its pure states to a system in state $\rho$ in terms of thermo-majorization curves. The partition function of the total system is $\left(1 + e^{-\beta W}\right) Z_S$ and the value of $W$ causes the curve of $\rho \otimes |1\rangle\langle 1|$ to stretch/compress with respect to the $x$-axis by a factor of $e^{-\beta W}$ relative to the curve of $\rho \otimes |0\rangle\langle 0|$ depending on whether $W$ is negative/positive. For a block-diagonal target state, $\sigma$, the work of transition is the value of $W$ which places the thermo-majorization curve of $\sigma \otimes |1\rangle\langle 1|$, just below the curve of $\rho \otimes |0\rangle\langle 0|$. Note that if $\sigma$ is not block-diagonal, it can occur that there is no value of $W$ for

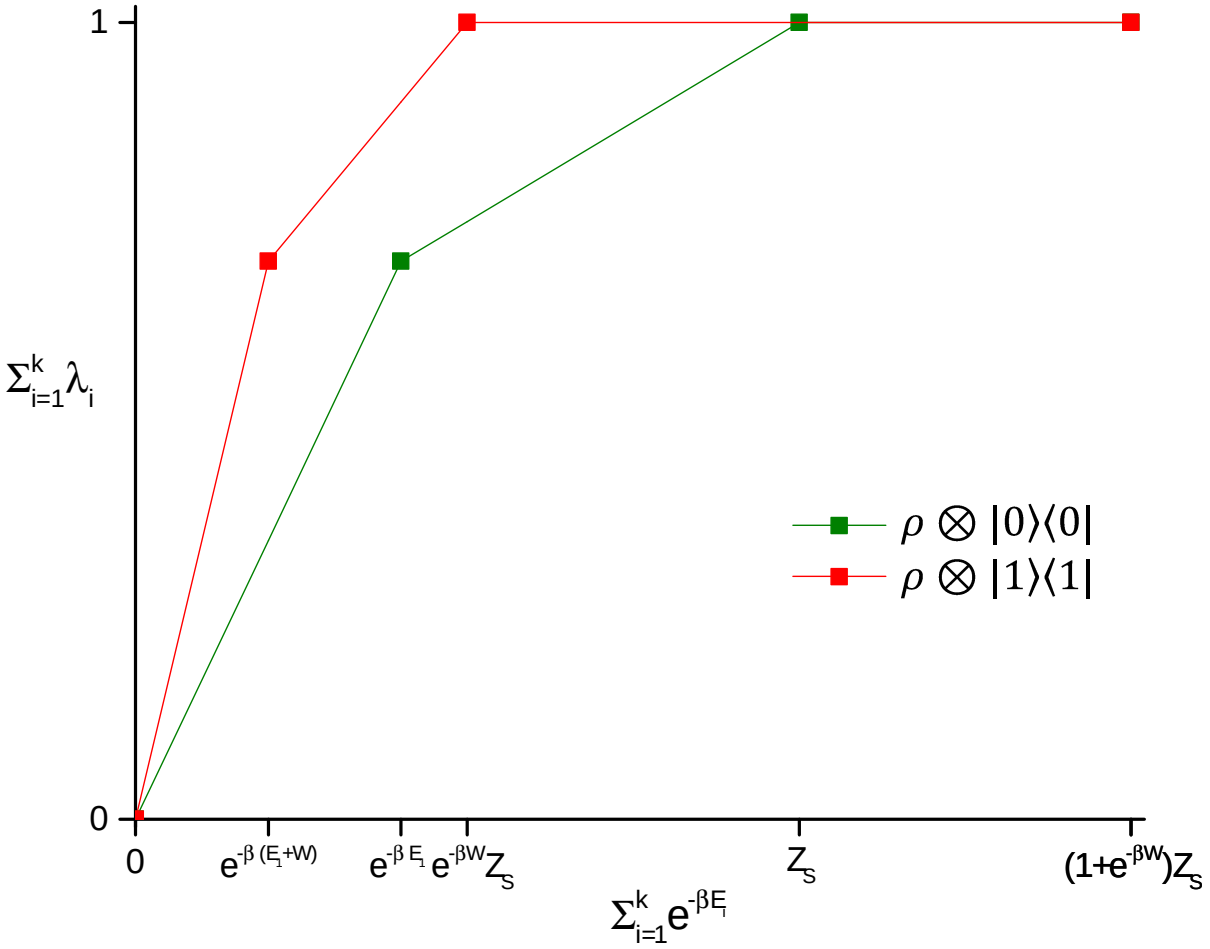



Figure 4.7: *Work and thermo-majorization curves.* Appending the non-zero energy state of the work storage system rescales the thermo-majorization curve of $\rho$ with respect to the $x$-axis. Here the value of $W$ is positive, compressing the curve.

which Eq. (4.29) holds. In these instances, an additional resource, a source of coherences, is needed to create $\sigma$ from $\rho$.

By setting $\sigma = \tau_S$, the thermal state of the system Hamiltonian, Eq. (4.29) provides the definition for the amount of *distillable work* from $\rho$. Similarly, by considering the transition from $\tau_S$ to $\rho$, the *work of formation* can be defined. These quantities can be expressed in terms of single-shot free energies, $F_{\min}$ and $F_{\max}$ [89] (see also [2, 61] for a derivation of these

expressions under an alternative sets of operations). The work of distillation is given by:

$$\begin{aligned} W_{\text{distil}}(\rho) &= F_{\min}(\rho) + k_B T \ln Z_S, \\ &= -k_B T \left[ \ln \sum_{i:\eta_i>0} e^{-\beta E_i} - \ln Z_S \right], \end{aligned} \tag{4.30}$$

while for block-diagonal $\rho$, the work of formation is given by:

$$\begin{aligned} W_{\text{form}}(\rho) &= -F_{\max}(\rho) - k_B T \ln Z_S, \\ &= -k_B T \left[ \ln \left( \eta_1^{(\rho)} e^{\beta E_1^{(\rho)}} \right) + \ln Z_S \right]. \end{aligned} \tag{4.31}$$

Note that the work of distillation is always non-negative, while the work of formation is never positive. For generic $\rho$, the absolute values for the distillable work and work of formation do not coincide and this implies that thermodynamics is irreversible at the nano-scale. This is in stark contrast to when the thermodynamic limit is taken. In this regime (allowing for some smoothing[3]) both $F_{\min}$ and $F_{\max}$ converge to the standard free energy as given in Eq. (4.1) and work distilled from $\rho$ can be used to recreate it.

In Chapter 5, we shall define the generic work of transition in terms of an optimization problem but as this optimization is closely related to the concept of a probabilistic transition, we defer doing this until Section 5.3.2.

### Changing Hamiltonian

Traditionally, thermodynamics is not just concerned with a system with Hamiltonian, $H$, and whether it is possible to transform $\rho$ into $\sigma$ whilst keeping $H$ fixed. One also wants to be able

[3] Suppose a state, $\rho$, has full rank. Then, by Eq. (4.30), no work can be deterministically distilled from it, even if one of the eigenvalues of $\rho$ is infinitesimally small. This is an undesirable feature as in an experiment one will never be able to precisely specify the state of the system. The solution, as discussed in [54] for noisy operations, is to define a smoothed version of the work of distillation, $W^{\epsilon}_{\text{distil}}$, where a small probability, $\epsilon$, of failing to draw work is allowed. It is shown in [89] that for thermal operations the correct notion of a soothed work of distillation is given by:

$$W^{\epsilon}_{\text{distil}}(\rho) = -k_B T \left[ \ln \left( \tilde{L}_{1-\epsilon}(\rho) \right) - \ln Z_S \right],$$

where $\tilde{L}_y$ denotes the horizontal distance between a state's thermo-majorization curve and the $y$-axis at $y$, a quantity we shall discuss more fully in Section 5.3.2.

Similarly, when considering the work of formation, one can define a smoothed version by considering:

$$W^{\epsilon}_{\text{form}}(\rho) = \max_{\rho'} W_{\text{form}}(\rho'),$$

where $\rho'$ is normalized and such that $D(\rho', \rho) \leq \epsilon$. Here, $D(\rho', \rho)$ denotes the trace distance between $\rho$ and $\rho'$.

to consider transitions where the Hamiltonian changes, i.e.:

$$(\rho, H_1) \longrightarrow (\sigma, H_2)\,, \tag{4.32}$$

and the work cost or yield of performing such a change. Using the *switch qubit* construction of [89], this scenario can be mapped to one with identical initial and final Hamiltonian if we instead consider the transition between $\rho \otimes |0\rangle\langle 0|$ and $\sigma \otimes |1\rangle\langle 1|$ with Hamiltonian:

$$H = H_1 \otimes |0\rangle\langle 0| + H_2 \otimes |1\rangle\langle 1|. \tag{4.33}$$

Note that the partition function associated with $H$ is $Z = Z_1 + Z_2$ where $Z_i$ is the partition function for Hamiltonian $H_i$.

The height of the thermo-majorization curve of $\rho \otimes |0\rangle\langle 0|$ with respect to $H$, is identical to that of $\rho$ with respect to $H_1$ on $[0, Z_1]$ and equal to 1 on $[Z_1, Z]$. Similarly, the height of the thermo-majorization curve of $\sigma \otimes |1\rangle\langle 1|$ is identical to that of $\sigma$ on $[0, Z_2]$ and equal to 1 on $[Z_2, Z]$. Hence, by extending the definition of $\tilde{V}_x(\rho)$ so that $\tilde{V}_x(\rho) = 1$ for $x \geq Z_1$, we can readily apply Theorem 17 to the case of changing Hamiltonian.

To model the work of transition in such a process, we combine this with Eq. (4.29) to give:

$$(\rho \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|, H + H_W) \xrightarrow{\text{TO}} (\sigma \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1|, H + H_W)\,, \tag{4.34}$$

and maximize over $W$.

## 4.3 Beyond thermo-majorization

Thermal operations are not the only paradigm for investigating thermodynamics outside of the thermodynamic limit and the description given here, though sufficient for the purposes of the remainder of this thesis, does not represent all that is known about them. We conclude this section by touching upon some other results and modifications to this framework.

Beyond this, a recent summary of progress made in applying tools and concepts from quantum information theory to thermodynamics can be found in [76]. In particular, Landauer's erasure principle has been expanded upon [56, 146, 65], the third law of thermodynamics has been derived and quantified in terms of the time it takes to cool quantum systems [121] and axiomatizations for defining work have been investigated [70]. Fluctuation relations, such as Crooks' theorem [51] and the Jarzynski equality [97], which relate statements about states in equilibrium to the results of non-equilibrium processes, have been compared to, and adjusted

in light of, single-shot results [79, 151, 53]. Small thermal machines and fridges [112] have been designed with the Carnot efficiency investigated both for these constructions, [156, 141], and in general [172].

### 4.3.1 Thermodynamics with catalysts

In a resource theory, the set of transitions that can be accomplished under the class of free operations can be expanded if one is allowed to make use of an additional system as a catalyst which is required to be returned in its initial state after the transformation. This occurs, for example, in pure state entanglement theory [99] and also impacts upon what is achievable under thermal operations [24]. In *catalytic thermal operations* (CTO), given $\rho$ and $\sigma$, we are interested in whether there exists a catalytic state, $\omega$, with associated Hamiltonian, $H_C$, such that:

$$(\rho \otimes \omega, H_S + H_C) \xrightarrow{\text{TO}} (\sigma \otimes \omega, H_S + H_C)\,. \tag{4.35}$$

If such an $\omega$ exists, we say $\rho \xrightarrow{\text{CTO}} \sigma$ and there exist instances for which $\rho \overset{\text{TO}}{\not\longrightarrow} \sigma$ and yet $\rho \xrightarrow{\text{CTO}} \sigma$. Investigating when such catalytic transformations occur has led to a family of 'second laws of thermodynamics' [24] that provide necessary and sufficient conditions for transitions to block-diagonal target states and necessary conditions otherwise. These laws can be stated in terms of generalized free energies, based on Rényi divergences, of which the standard free energy, Eq. (4.1), and the single-shot free energies, Eqs. (4.30) and (4.31), are particular instances.

How do these second laws change if, rather than demanding that $\omega$ is returned exactly, we instead ask only for a state, $\omega'$, close to $\omega$ with respect to some distance? Three regimes are considered in [24]. Firstly, if the requirement is only that $\omega'$ is $\epsilon$-close to $\omega$ in trace distance, for fixed $\epsilon$, then by choosing $\omega$ to have a sufficiently large dimension, it is always possible to convert any $\rho$ into any $\sigma$ with no restriction. This is the thermal operations equivalent of entanglement embezzling [163]. If the demand is instead such that the above $\epsilon$ scales like $O\left(\frac{1}{\log n_C}\right)$, where $n_C$ is the dimension of the catalyst system, then the set of second laws collapse to the standard free energy of Eq. (4.1). However, the full set is recovered if we require $\omega'$ and $\omega$ to be close in the sense that $W_{\omega' \to \omega} \leq \epsilon$, for fixed $\epsilon$.

One further way in which catalysts can be abused, is by allowing them to become correlated. In [117] it has been shown that the transformation:

$$\rho \otimes \omega^1_{C_1} \otimes \omega^2_{C_2} \otimes \omega^3_{C_3} \xrightarrow{\text{TO}} \sigma \otimes \Omega_{C_1C_2C_3}, \tag{4.36}$$

where $\mathrm{Tr}_{\setminus i}\left[\Omega_{C_1C_2C_3}\right] = \omega^i_{C_i}$, for all $i$, can be performed if and only if $F(\rho) \geq F(\sigma)$. This surprising result is somewhat counter-intuitive as one would expect that creating correlations between the catalysts should cost work.

### 4.3.2 Dealing with coherences

As we have been careful to note in this chapter, criteria derived from thermo-majorization are generally only sufficient for characterizing if a transition is allowed when the target state is block-diagonal in the energy eigenbasis. It can therefore be argued, that such results are not fully quantum in nature and to be regarded as such, they need to be generalized to arbitrary target states. Towards this end, allowed transitions have been completely characterized for qubits [52] and additional necessary conditions, independent of the generalized free energies, have been found [115, 116].

In particular, for quantum systems, work can be 'locked' inside states with coherences [115]. For example, from the pure state $|\tau\rangle$, defined such that $|\tau\rangle\langle\tau|_D = \tau$ (the thermal state of the system), no work can be extracted using thermal operations. To do so, a reference state with sufficient coherence must be used [115]. This additional state can however, be chosen in such a way for it to almost be regarded as a catalyst [3]. While its state changes during the process, so it is not a true catalyst, by expending some work after each use it is possible to preserve the system's ability to facilitate future extractions. By choosing the dimension of this coherence catalyst to be large, one can make the probability of extracting positive work from $|\tau\rangle$ (after paying the addition cost) close to one [105].

### 4.3.3 Average energy conservation

As an alternative to thermal operations, rather than demanding that the allowed unitaries commute with the total Hamiltonian and considering single-shot work extraction, one could instead require that energy is preserved only on average and analyze the average work that is extracted. In this paradigm, the average work that can be extracted in converting one state into another is given by the difference of the standard free energy, regardless of whether either state contains coherences [157]. If one considers average work extraction with respect to energy conserving unitaries, it is also shown that it is possible to extract the free energy difference when the states are not coherent.

Average energy conservation is however, not as appealing as the strict energy conservation

of thermal operations from an operational viewpoint as the set of allowed operations depend on the initial state being manipulated. As such, they are harder to formulate mathematically than thermal operations and a machine built to implement them will be dependent on the state supplied as an input.

## 4.4 Summary

This chapter has introduced the resource theory approach to nano-scale thermodynamics as formalized by thermal operations. As an illustrative sandbox for testing ideas, the theory of noisy operations was also discussed and this can be regarded as modeling thermodynamics in the absence of energy.

When we do not take the thermodynamic limit, determining whether a transition can occur is given by majorization criteria such as thermo-majorization. These can be neatly illustrated using Lorenz and thermo-majorization curves which enable one to determine how much work can be extracted from a state, how much work is required to form it and what the work cost/yield of a given transformation is. Suppose our initial state does not thermo-majorize the target state. Can anything be achieved without supplying any additional work? In Chapter 5 we will define the notion of a probabilistic transition to work towards addressing this question.

Despite the progress highlighted here, there are still many open questions regarding thermal operations. As summarized in Section 4.3.2, many of these revolve around the role of coherence within the theory and how much of it can be regarded as truly quantum. Furthermore, the thermo-majorization criteria derived here make assumptions on the structure of the heat bath. In particular, it is assumed that the degeneracies of the energy levels in the heat bath grow exponentially with energy. How do the laws of thermodynamics look without this ideal bath? Finally, the processes allowed in thermal operations assume that an experimenter has extremely fine-tuned control over the degrees of freedom in both the system and bath. While this means that the thermo-majorization constraints hold even with imperfect control, are there other restrictions on what can actually be achieved at the nano-scale? This will be the subject of Chapter 6.

# Chapter 5

# Probabilistic Thermodynamical Transitions

## 5.1 Probabilistic transitions

Given two states, $\rho$ and $\sigma$, where $\sigma$ is block-diagonal in the energy eigenbasis, Theorem 16 tells us that it is possible to transform $\rho$ into $\sigma$ using thermal operations if and only if $\sigma$ is thermo-majorized by $\rho$. If $\sigma$ is not thermo-majorized by $\rho$, the transition is still possible, provided sufficient work is provided and one can compute the work required (or gained) from this transition using thermo-majorization curves [89] or via the relative-mixedness defined in [61].

Suppose however, that we want to make a transition from $\rho$ to $\sigma$, and it requires work that we cannot, or do not wish to, expend. Can we still make the transition albeit with some probability, $p$, rather than with certainty? If so, what is the highest probability, $p^*$, that can be achieved? More specifically, given $\rho$ and $\sigma$, we are interested in maximizing $p$ in the following process:

$$\rho \xrightarrow{\text{TO}} \rho' = p\sigma + (1-p)\, X, \tag{5.1}$$

with $X$ being some arbitrary state. Furthermore, given $p^*$, does there exist a measurement one can perform on $\rho'$ such that one obtains $\sigma$ with that probability? Care has to be taken when answering this last question. Measurements do not come completely for free in thermodynamics - as discussed for the Szilard engine in Section 4.1, it costs work to erase the record of the measurement outcome [109, 20]. As such we define what we mean by a probabilistic thermo-

dynamical transition by Eq. (5.1) and will only allow a measurement after such a process has taken place.

## 5.2 Noisy operations

Before investigating Eq. (5.1) in the context of thermal operations, we will first consider the simpler, special case of noisy operations.

### 5.2.1 Non-deterministic transitions

Under noisy operations, $\rho$ can be transformed into $\sigma$ with certainty if and only if $\rho$ majorizes $\sigma$ (Theorem 15). Here we shall transform $\rho$ into $\sigma$ probabilistically and determine the maximum probability with which it can be achieved. A similar problem was considered in [164] for pure state entanglement manipulation and adapting its techniques can be used to show:

**Theorem 18.** *Suppose we wish to transform the state of an $n$-level system from $\rho$ into $\sigma$ under noisy operations. The maximum value of $p$ that can be achieved in the transition:*

$$\rho \xrightarrow{NO} \rho' = p\sigma + (1-p)\, X, \tag{5.2}$$

*is given by:*

$$p^* = \min_{l \in \{1,\ldots,n\}} \frac{V_l(\rho)}{V_l(\sigma)}, \tag{5.3}$$

*where $V_l$ are the monotones defined in Eq.* (4.9).

*Proof.* The proof is split into two parts: first we show that it is impossible to achieve a value of $p$ greater than that in Eq. (5.3) and then we give a protocol obtaining $p = p^*$. Let $\vec{\eta} = \{\eta_1, \ldots \eta_n\}$, $\vec{\eta'} = \{\eta'_1, \ldots \eta'_n\}$ and $\vec{\zeta} = \{\zeta_1, \ldots \zeta_n\}$ denote the ordered eigenvalues of $\rho$, $\rho'$ and $\sigma$.

Towards our first goal, we begin by showing that given Eq. (5.2):

$$V_l(\rho) \geq pV_l(\sigma), \quad \forall l. \tag{5.4}$$

To see this note that from Weyl's inequality [168, 87], we have:

$$\eta'_i \geq p\zeta_i + (1-p)\, x_n, \quad \forall i,$$

where $x_n$ is the smallest eigenvalue of $X$. As $X$ is a positive semidefinite matrix, $x_n \geq 0$ and:

$$\eta'_i \geq p\zeta_i, \quad \forall i. \tag{5.5}$$

Hence:

$$V_l(\rho) \geq V_l(\rho') = \sum_{i=1}^{l} \eta_i' \geq p \sum_{i=1}^{l} \zeta_i = pV_l(\sigma),$$

where the first inequality uses Eq. (4.11) together with the fact that $\rho$ majorizes $\rho'$ and the second follows from Eq. (5.5).

Now suppose it was possible to achieve a value of $p$ greater than $p^*$ in Eq. (5.2). Then there would exist an $l$ such that $V_l(\rho) < pV_l(\sigma)$, contradicting Eq. (5.4).

To show that $p^*$ is obtainable, we define the following quantities. First, define $l_1$ by:

$$l_1 = \max\left\{ l : \frac{V_l(\rho)}{V_l(\sigma)} = p^* \equiv r^{(1)} \right\}.$$

Then we proceed iteratively and, provided $l_{i-1} < n$, define:

$$r^{(i)} = \min_{l > l_{i-1}} \frac{V_l(\rho) - V_{l_{i-1}}(\rho)}{V_l(\sigma) - V_{l_{i-1}}(\sigma)},$$

so, noting that $V_l(\rho) - V_{l_{i-1}}(\rho) = \sum_{j=l_{i-1}+1}^{l} \eta_j$ for $l > l_{i-1}$ (and that a similar expression holds for $\sigma$), we have:

$$r^{(i)} \sum_{j=l_{i-1}+1}^{l} \zeta_j \leq \sum_{j=l_{i-1}+1}^{l} \eta_j, \quad \forall l > l_{i-1}. \tag{5.6}$$

Define $l_i$ by:

$$l_i = \max\left\{ l : l > l_{i-1}, \frac{V_l(\rho) - V_{l_{i-1}}(\rho)}{V_l(\sigma) - V_{l_{i-1}}(\sigma)} = r^{(i)} \right\}.$$

Note that we have $r^{(i)} > r^{(i-1)}$. To see this, first observe that for $a, b, c, d > 0$:

$$\frac{a}{b} < \frac{a+c}{b+d} \Leftrightarrow \frac{a}{b} < \frac{c}{d}. \tag{5.7}$$

Setting:

$$\begin{aligned} a &= V_{l_{i-1}}(\rho) - V_{l_{i-2}}(\rho), \\ b &= V_{l_{i-1}}(\sigma) - V_{l_{i-2}}(\sigma), \\ c &= V_{l_i}(\rho) - V_{l_{i-1}}(\rho), \\ d &= V_{l_i}(\sigma) - V_{l_{i-1}}(\sigma), \end{aligned}$$

so $\frac{a}{b} = r^{(i-1)}$ and $\frac{c}{d} = r^{(i)}$, then:

$$\frac{a+c}{b+d} = \frac{V_{l_i}(\rho) - V_{l_{i-2}}(\rho)}{V_{l_i}(\sigma) - V_{l_{i-2}}(\sigma)} > r^{(i-1)} = \frac{a}{b},$$

where the inequality follows from the definition of $r^{(i-1)}$. Using Eq. (5.7), the claim that $r^{(i)} > r^{(i-1)}$ now follows. Overall, this protocol generates a set of $l_i$ such that $0 = l_0 < l_1 < \ldots < l_k = n$ and a set of $r_i$ such that $p^* = r^{(1)} < \ldots < r^{(k)}$.

Now we split $\rho$ and $\sigma$ into blocks and define:

$$\rho_i = \mathrm{diag}\left(\eta_{l_{i-1}+1}, \ldots, \eta_{l_i}\right),$$
$$\sigma_i = \mathrm{diag}\left(\zeta_{l_{i-1}+1}, \ldots, \zeta_{l_i}\right).$$

Then from Eq. (5.6) (and the fact that equality occurs when $l = l_i$), we have that $\rho_i$ majorizes $r^{(i)}\sigma_i$ and we can perform:

$$\rho_i \xrightarrow{NO} r^{(i)}\sigma_i = p^*\sigma_i + \left(r^{(i)} - p^*\right)\sigma_i, \quad \forall i. \tag{5.8}$$

With a bit of massaging and recombining the blocks, this is the same form as Eq. (5.2) with $p = p^*$ and with the blocks of $X$ being defined by:

$$X_i = \frac{r^{(i)} - p^*}{1 - p^*}\sigma_i.$$

□

Note that as $V_n(\rho) = V_n(\sigma) = 1$ and $\eta_1 > 0$, we are guaranteed that $0 < p^* \leq 1$.

If we want to obtain $\sigma$ from $\rho$ with probability $p^*$ rather than leaving it as part of a probabilistic mixture as per Eq. (5.2), we can do so by performing a two outcome measurement, with measurement operators $\left\{\sqrt{M}, \sqrt{\mathbb{I} - M}\right\}$, where the blocks of $M$ are given by:

$$M_i = \mathrm{diag}\left(\frac{p^*}{r^{(i)}}, \ldots, \frac{p^*}{r^{(i)}}\right). \tag{5.9}$$

To see that this defines a valid measurement, note that $0 \leq \frac{p^*}{r^{(i)}} \leq 1$, for all $i$. After applying this measurement to $\rho'$ and reading the result, we will produce either:

$$\sqrt{M}\,\rho'\sqrt{M^\dagger} = p^*\sigma,$$

or

$$\sqrt{(\mathbb{I} - M)}\,\rho'\sqrt{(\mathbb{I} - M)^\dagger} = (1 - p^*)\,X,$$

and hence we obtain $\sigma$ with probability $p^*$.

However, performing this measurement is outside of the class of noisy operations and hence costs work. As such, if a general two outcome measurement is allowed at any point during the

protocol, it can be possible to transform $\rho$ into $\sigma$ with probability greater than $p^*$. For example, if $\rho$ and $\sigma$ are qubits, we can convert $\rho$ into $\sigma$ with certainty using this extra resource. Firstly, we add an additional qubit in the maximally mixed state and measure it in the computational basis. This results in a pure state, either $|0\rangle$ or $|1\rangle$. As these majorize all other qubit states, we can use it to obtain any $\sigma$ with certainty. It is for this reason that we restrict ourselves to optimizing $p^*$ in Eq. (5.1) and demand that any measurement must occur after the transformation has taken place and with no further processing.

### 5.2.2 Quantifying the purity of transition

The maximum probability of a transition under noisy operations given by Eq. (5.3) is written as a ratio of monotones of the theory. These monotones can be defined in terms of the Lorenz curve of a state as the function $V_l(\rho)$ is equal to the height of the Lorenz curve of $\rho$ at $x = \frac{l}{n}$. They are however, not the only monotones that can be constructed from a Lorenz curve. An alternative set, $L_y(\rho)$ where $0 \leq y \leq 1$, can be defined as the shortest horizontal distance between the Lorenz curve of $\rho$ and the $y$-axis at $y$. Note that these functions *never decrease* under noisy operations and in particular:

$$\begin{aligned} L_{y_k}(\rho) &= \frac{k}{n}, \qquad \text{for } y_k = \sum_{i=1}^{k} \eta_i, \qquad 1 \leq k < \operatorname{rank}(\rho), \\ L_1(\rho) &= \frac{\operatorname{rank}(\rho)}{n}, \end{aligned} \tag{5.10}$$

where $\vec{\eta} = \{\eta_1, \ldots, \eta_n\}$ are again the ordered eigenvalues of $\rho$.

If we define the set $\mathcal{D}(\rho)$ by:

$$\mathcal{D}(\rho) = \left\{ \sum_{i=1}^{k} \eta_i \right\}_{k=1}^{\operatorname{rank}(\rho)}, \tag{5.11}$$

Then a transition from $\rho$ to $\sigma$ is achievable with certainty under noisy operations if and only if:

$$L_y(\rho) \leq L_y(\sigma), \quad \forall y \in \mathcal{D}(\sigma). \tag{5.12}$$

That it is sufficient to consider only those $y \in \mathcal{D}(\sigma)$ will be justified below.

These horizontal monotones, $L_y$, also allow us to quantify the optimal purity of transition that is required or extracted in converting $\rho$ into $\sigma$:

**Lemma 13.** *Given two states, $\rho$ and $\sigma$, the purity that can be extracted or is required in transforming $\rho$ into $\sigma$ under noisy operations, $S_{\rho\to\sigma}$, is given by:*

$$2^{-S_{\rho\to\sigma}} = \max_{y\in\mathcal{D}(\sigma)} \frac{L_y(\rho)}{L_y(\sigma)}. \tag{5.13}$$

*Proof.* Note first that:

$$2^{-S_{\rho\to\sigma}} = \max_{y\in[0,1]} \frac{L_y(\rho)}{L_y(\sigma)}. \tag{5.14}$$

This follows from the fact that to obtain the optimal value of $S_{\rho\to\sigma}$, we wish to rescale the Lorenz curve of $\rho$ with respect to the $x$-axis in such a way that it just majorizes that of $\sigma$. The curves should touch but not cross. The amount that we need to rescale by is given by Eq. (5.14).

We now show that it is sufficient to maximize over $y \in \mathcal{D}(\sigma)$. Let $s_0 = 0$ and $s_k = \sum_{i=1}^{k} \zeta_i$ for $k \in \{1, \ldots, \operatorname{rank}(\sigma)\}$. Then, for $j \in \{1, \ldots, \operatorname{rank}(\sigma)\}$, as the Lorenz curve of $\sigma$ is a straight line on the interval $[s_{j-1}, s_j]$ and the Lorenz curve of $\rho$ is concave:

$$\max_{y\in[s_{j-1},s_j]} \frac{L_y(\rho)}{L_y(\sigma)} \leq \max_{r\in[0,1]} \frac{rL_{s_{j-1}}(\rho) + (1-r)L_{s_j}(\rho)}{r\frac{j-1}{n} + (1-r)\frac{j}{n}}. \tag{5.15}$$

It is straightforward to check that the maximum value occurs at either $r = 0$ or $r = 1$. We can thus replace the inequality in Eq. (5.15) with an equality and it follows that it suffices to maximize over $y \in \mathcal{D}(\sigma)$. □

As $\rho \xrightarrow{\text{NO}}$ is possible if and only if $S_{\rho\to\sigma} \geq 0$, the finite set in Eq. (5.12) is justified.

Note that it was shown in [77] that it is possible to calculate $S_{\rho\to\sigma}$ by performing an optimization over the ratios calculated at the 'elbows' of both $\rho$ and $\sigma$. In Lemma 13 we have shown that it suffices to consider just the 'elbows' of $\sigma$.

### Bounds on the transition probability

The expression for $S_{\rho\to\sigma}$ given in Lemma 13 is strikingly similar to that for $p^*$ from Theorem 18. Both are calculated in terms of an optimization over ratios of monotones that can be defined from Lorenz curves. Here we shall show that this similarity can be used to bound $p^*$ in terms of $S_{\rho\to\sigma}$ and $S_{\sigma\to\rho}$:

**Lemma 14.** *Given two states, $\rho$ and $\sigma$, such that $S_{\rho\to\sigma} \leq 0$, then under noisy operations:*

$$2^{S_{\rho\to\sigma}} \leq p^* \leq 2^{-S_{\sigma\to\rho}}. \tag{5.16}$$

*If $S_{\rho\to\sigma} \geq 0$, $p^* = 1$ and the transformation from $\rho$ to $\sigma$ can be done deterministically, potentially extracting a finite amount of purity.*

*Proof.* We start with the lower bound, providing a protocol which achieves $p = 2^{S_{\rho\to\sigma}}$. Assuming $|S_{\rho\to\sigma}| = \log_2 \frac{d}{j}$ for simplicity, this protocol runs as follows:

$$\begin{aligned}
\rho &\xrightarrow{NO} \rho \otimes I_d, \\
&= \frac{j}{d}\rho \otimes s_{\log_2 \frac{d}{j}} + \frac{d-j}{d}\rho \otimes s_{\log_2 \frac{d}{d-j}}, \\
&\xrightarrow{NO} \frac{j}{d}\sigma \otimes I_d + \frac{d-j}{d}Y, \\
&\xrightarrow{NO} \frac{j}{d}\sigma + \frac{d-j}{d}\mathrm{Tr}_A Y,
\end{aligned}$$

where $A$ labels the ancilla system appended in the first noisy operation and $Y$ is the state obtained by applying the second noisy operation to $\rho \otimes s_{\log_2 \frac{d}{d-j}}$. Using this protocol, we obtain something of the form Eq. (5.2) with $p = 2^{S_{\rho\to\sigma}}$ and $X = \mathrm{Tr}_A Y$. As $p^*$ is the maximum value of $p$ obtainable in Eq. (5.2), we derive the lower bound.

We now consider the upper bound and to obtain a useful bound, assume $S_{\sigma\to\rho} > 0$. Recalling that $S_{\mathrm{form}}(\rho)$ denotes the minimum amount of purity required to form $\rho$, let $s_{S_{\mathrm{form}}(\rho)}$ be the least sharp state that majorizes $\rho$ (see Figure 4.3). Note that $S_{\mathrm{form}}$ decreases under noisy operations and is additive across tensor products [77]. In terms of the largest eigenvalues of $\rho$ and $\sigma$:

$$\begin{aligned}
s_{S_{\mathrm{form}}(\rho)} &= s_{\log_2(\eta_1 n)}, \\
s_{S_{\mathrm{form}}(\sigma)} &= s_{\log_2(\zeta_1 n)}.
\end{aligned}$$

By definition, as $S_{\sigma\to\rho} > 0$:

$$\sigma \xrightarrow{\mathrm{NO}} \rho \otimes s_{S_{\sigma\to\rho}}.$$

Now, using first the monotonicity of $S_{\text{form}}$ and then the additivity:

$$\begin{aligned}
S_{\text{form}}(\sigma) &\geq S_{\text{form}}\left(\rho \otimes s_{S_{\sigma\to\rho}}\right), \quad \text{(monotonicity)} \\
&= S_{\text{form}}(\rho) + S_{\sigma\to\rho}. \quad \text{(additivity)} \\
\Rightarrow S_{\sigma\to\rho} &\leq S_{\text{form}}(\sigma) - S_{\text{form}}(\rho), \\
&= \log_2(\zeta_1 n) - \log_2(\eta_1 n), \\
&= \log_2\left(\frac{\zeta_1}{\eta_1}\right). \\
\Rightarrow 2^{-S_{\sigma\to\rho}} &\geq \frac{\eta_1}{\zeta_1}, \\
&= \frac{V_1(\rho)}{V_1(\sigma)}, \\
&\geq p^*, \quad \text{(by definition)},
\end{aligned}$$

as required. □

Before moving on to consider probabilistic transitions and their relation to the work of transition in thermal operations, we note that the horizontal monotones defined here can equally be applied in entanglement theory.

### 5.2.3 Aside: Entanglement of transition under LOCC

The monotones that we have used for studying noisy operations, have been, or can be, defined solely in terms of Lorenz curves. They are also monotones in the resource theory of bipartite pure state manipulation under LOCC [133, 165], where such curves can also be constructed. Using our monotones, and the behavior of Lorenz curves under tensor product with certain states, we give an expression for the single-shot *entanglement of transition*. This is the amount of entanglement that must be added (or can be extracted) in transforming one pure state, $|\Psi_{AB}\rangle$, into another, $|\Phi_{AB}\rangle$, under LOCC.

Previous work has considered the *distillable entanglement* and *entanglement cost* - the entanglement of transition when one of $|\Phi_{AB}\rangle$ or $|\Psi_{AB}\rangle$, respectively, is taken to be a separable state. In [36], the amount of entanglement that can be distilled from a single copy of a bipartite mixed state, $\sigma_{AB}$, was bounded in terms of the coherent information. For a bipartite pure state, $|\Psi_{AB}\rangle$, it is given precisely by the min-entropy of the reduced state $\text{Tr}_B|\Psi_{AB}\rangle\langle\Psi_{AB}|$ [37]. The amount of entanglement required to create a single copy of $\sigma_{AB}$ was calculated in [38] in terms of the conditional zero-Rényi entropy. In each paper, the analysis extends to accomplishing

the task up to fixed error, $\epsilon$. Here we go beyond the distillation and cost, showing that the more general entanglement of transition between two arbitrary pure bipartite states, can be quantified in terms of the monotones $L_y$.

For a bipartite pure state, $|\Psi_{AB}\rangle$, on a system $AB$, let:

$$\rho_{|\Psi\rangle} = \mathrm{Tr}_B |\Psi_{AB}\rangle\langle\Psi_{AB}|. \tag{5.17}$$

Without access to any additional resources, it is possible for two separated parties to transform $|\Psi_{AB}\rangle$ into another bipartite state, $|\Phi_{AB}\rangle$, under LOCC if and only if $\rho_{|\Phi\rangle}$ majorizes $\rho_{|\Psi\rangle}$ [133]. Hence, if $|\Psi_{AB}\rangle$ can be transformed into $|\Phi_{AB}\rangle$:

$$V_l\left(\rho_{|\Phi\rangle}\right) \geq V_l\left(\rho_{|\Psi\rangle}\right), \quad \forall l, \tag{5.18}$$

and:

$$L_y\left(\rho_{|\Phi\rangle}\right) \leq L_y\left(\rho_{|\Psi\rangle}\right), \quad \forall y \in \mathcal{D}\left(\rho_{|\Psi\rangle}\right), \tag{5.19}$$

where the functions $V_l$, $L_y$ and the set $\mathcal{D}$ are defined as per Eqs. (4.9), (5.10) and (5.11) respectively. Note that for LOCC we consider the 'elbows' of the Lorenz curve associated with the initial state whilst for noisy operations we consider the 'elbows' of the final state's curve when determining if a transition is possible. This change occurs as for a transition to take place in pure state entanglement theory, we require that the final state majorizes the initial state whilst in the theory of noisy operations, we require that the initial state majorizes the final.

The unit for quantifying entanglement costs is the ebit - the maximally entangled state with local dimension 2. The maximally entangled state with local dimension $d$:

$$|e_d\rangle = \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle_A |i\rangle_B, \tag{5.20}$$

requires the two parties to share $\log_2 d$ ebits to prepare it and they can extract $\log_2 d$ shared ebits if they share one. Separable states are free within this resource theory so, if we define:

$$|\mathrm{sep}_d\rangle = |0\rangle_A |0\rangle_B, \tag{5.21}$$

as a separable pure state with local dimension $d$, then $|\mathrm{sep}_d\rangle$ costs 0 ebits to prepare and no shared entanglement can be extracted from it. Note that:

$$L_y\left(\rho_{|\Psi\rangle\otimes|e_d\rangle}\right) = L_y\left(\rho_{|\Psi\rangle}\right), \tag{5.22}$$

$$L_y\left(\rho_{|\Psi\rangle\otimes|\mathrm{sep}_d\rangle}\right) = \frac{1}{d} L_y\left(\rho_{|\Psi\rangle}\right). \tag{5.23}$$

The entanglement of transition, $E_{|\Psi\rangle\to|\Phi\rangle}$, is the optimal amount of shared, bipartite entanglement that the parties need to add, or can gain, to transform a copy of $|\Psi_{AB}\rangle$ into $|\Phi_{AB}\rangle$ under LOCC. If the quantity is negative, entanglement must be used up to make the transition possible while if it is positive, entanglement can be extracted. $E_{|\Psi\rangle\to|\Phi\rangle}$ is the maximum value of $v\log_2 d_2 - u\log_2 d_1$ that can be achieved where $u, v, d_1, d_2 \in \mathbb{N}$ are such that:

$$|\Psi\rangle|e_{d_1}\rangle^{\otimes u}|\mathrm{sep}_{d_2}\rangle^{\otimes v} \overset{LOCC}{\longrightarrow} |\Phi\rangle|e_{d_2}\rangle^{\otimes v}|\mathrm{sep}_{d_1}\rangle^{\otimes u}. \tag{5.24}$$

In terms of Lorenz curves, the addition of these entangled and separable state serve to rescale (with respect to the $x$-axis) the curves associated with $|\Psi_{AB}\rangle$ and $|\Phi_{AB}\rangle$ by ${d_2}^{-v}$ and ${d_1}^{-u}$ respectively. To maximize $E_{|\Psi\rangle\to|\Phi\rangle}$, the Lorenz curve of the rescaled $|\Psi_{AB}\rangle$ needs to lie just to the right of the Lorenz curve of the rescaled $|\Phi_{AB}\rangle$. Hence:

$$\frac{1}{{d_2}^v} L_y\left(\rho_{|\Psi\rangle}\right) \geq \frac{1}{{d_1}^u} L_y\left(\rho_{|\Phi\rangle}\right), \quad \forall y \in \mathcal{D}\left(\rho_{|\Psi\rangle}\right), \tag{5.25}$$

with equality for some $y$. This gives:

$$2^{-\left(E_{|\Psi\rangle\to|\Phi\rangle}\right)} = \frac{{d_1}^u}{{d_2}^v} = \max_{y\in\mathcal{D}\left(\rho_{|\Psi\rangle}\right)} \frac{L_y\left(\rho_{|\Phi\rangle}\right)}{L_y\left(\rho_{|\Psi\rangle}\right)}, \tag{5.26}$$

in analogy with Lemma 13 for the work of transition in noisy operations.

This can be generalized to consider situations where we require only that the final state is $\epsilon$-close to the target state $|\Phi\rangle$ with respect to a measure such as the squared fidelity, $F^2\left(|\Phi'\rangle, |\Phi\rangle\right) = |\langle\Phi'|\Phi\rangle|^2$. Let:

$$b^\epsilon\left(|\Phi\rangle\right) = \left\{|\Phi'\rangle : \left|\langle\Phi'|\Phi\rangle\right|^2 \geq 1-\epsilon\right\}. \tag{5.27}$$

Then, defining $E^\epsilon_{|\Psi\rangle\to|\Phi\rangle}$ by:

$$E^\epsilon_{|\Psi\rangle\to|\Phi\rangle} = \max_{|\Phi'\rangle\in b^\epsilon(|\Phi\rangle)} E_{|\Psi\rangle\to|\Phi'\rangle}, \tag{5.28}$$

we can write:

$$E^\epsilon_{|\Psi\rangle\to|\Phi\rangle} = \max_{|\Phi'\rangle\in b^\epsilon(|\Phi\rangle)} \left\{-\log_2\left[\max_{y\in\mathcal{D}\left(\rho_{|\Psi\rangle}\right)} \frac{L_y\left(\rho_{|\Phi'\rangle}\right)}{L_y\left(\rho_{|\Psi\rangle}\right)}\right]\right\}. \tag{5.29}$$

## 5.3 Thermal operations

Using the results of Section 5.2 as guidance, we now turn to quantifying the maximum probability of a thermodynamical transition in full-blown thermal operations. We derive the results assuming that the initial and target states are associated with the same Hamiltonian but they can readily be extended to the case of changing Hamiltonian using the construction given in Eq. (4.33).

### 5.3.1 Non-deterministic transitions

Recall from Theorems 16 and 17 that $\rho$ can be deterministically converted into a block-diagonal $\sigma$ under thermal operations if and only if $\rho$ thermo-majorizes $\sigma$ and that this can be formulated in terms of the monotones $\tilde{V}_x$, the height of the thermo-majorization curve at $x$. These can be used to bound the maximum probability with which a probabilistic transition can happen and when $\sigma$ is block-diagonal, this bound can be achieved:

**Theorem 19.** *Suppose we wish to transform the state $\rho$ into the state $\sigma$ under thermal operations for an $n$-level system with Hamiltonian $H = \sum_{i=1}^{n} E_i|i\rangle\langle i|$. The maximum value of $p$, $p^*$, that can be achieved in the transition:*

$$\rho \xrightarrow{TO} \rho' = p\sigma + (1-p)\,X, \tag{5.30}$$

*is such that:*

$$p^* \leq \min_{x \in \mathcal{L}(\sigma)} \frac{\tilde{V}_x(\rho)}{\tilde{V}_x(\sigma)}. \tag{5.31}$$

*Furthermore, if $\sigma$ is block-diagonal in the energy eigenbasis, there exists a protocol that achieves the bound.*

*Proof.* Proving this result is more complicated than proving Theorem 18 due to the fact that $\rho$ and $\sigma$ may have different $\beta$-orderings. We proceed as before, first showing the bound in Eq. (5.31) and then giving a protocol that achieves the bound when $\sigma$ is block-diagonal.

We begin by showing that given Eq. (5.30):

$$\tilde{V}_x(\rho) \geq p\tilde{V}_x(\sigma), \quad \forall x.$$

First consider (for general $\sigma$) the maximum value of $p$ that can be achieved in attempting to convert $\rho$ into $\sigma$. As decohering is a thermal operation, this value of $p$ can also be achieved when attempting to convert $\rho$ into $\sigma_D$:

$$\begin{aligned} \rho &\xrightarrow{TO} \rho' &&= p\sigma + (1-p)\,X, \\ &\xrightarrow{decohere} \rho'_D &&= p\sigma_D + (1-p)\,X_D. \end{aligned}$$

Thus, to upper bound $p^*$, it suffices to show that Eq. (5.31) holds for block-diagonal $\sigma$. Furthermore, without loss of generality, we can assume that $\rho'$ and $X$ are also block-diagonal. Using Weyl's inequality as per Theorem 18 to deal with degenerate energy levels, for block-diagonal $\rho'$, $\sigma$ and $X$, we have:

$$\eta'_i \geq p\zeta_i, \quad \forall i. \tag{5.32}$$

where $\{\eta'_i\}_{i=1}^n$ and $\{\zeta_i\}_{i=1}^n$ are the (unordered) eigenvalues of $\rho'$ and $\sigma$.

Now consider the sub-normalized thermo-majorization curve of $p\sigma$ given by the points:

$$\left\{\left(\sum_{i=1}^{k} e^{-\beta E_i^{(\sigma)}}, p\sum_{i=1}^{k} \zeta_i^{(\sigma)}\right)\right\}_{k=1}^{n}, \tag{5.33}$$

and the (possibly non-concave) curve formed by plotting the eigenvalues of $\rho'$ according to the $\beta$-ordering of $\sigma$. This is given by the points:

$$\left\{\left(\sum_{i=1}^{k} e^{-\beta E_i^{(\sigma)}}, \sum_{i=1}^{k} \eta_i'^{(\sigma)}\right)\right\}_{k=1}^{n}. \tag{5.34}$$

By Eq. (5.32), the curve defined in Eq. (5.34) is never below that defined in Eq. (5.33).

Finally, the thermo-majorization curve of $\rho'$ is given by:

$$\left\{\left(\sum_{i=1}^{k} e^{-\beta E_i^{(\rho')}}, \sum_{i=1}^{k} \eta_i'^{(\rho')}\right)\right\}_{k=1}^{n}. \tag{5.35}$$

Note that attempting to construct a thermo-majorization curve for $\rho'$ with respect to the $\beta$-ordering of another state, as we do in Eq. (5.34), has the effect of rearranging the piecewise linear segments of the true thermo-majorization curve. This means that they may no longer be joined from left to right in order of decreasing gradient. Such a curve will always be below the true thermo-majorization curve. To see this, imagine constructing a curve from the piecewise linear elements and in particular, trying to construct a curve that would lie above all other possible constructions. Starting at the origin, we are forced to choose the element with the steepest gradient - all other choices would lie below this by virtue of having a shallower gradient. We then proceed iteratively, starting from the endpoint of the previous section added and choosing the element with the largest gradient from the remaining linear segments. The construction that we obtain is the true thermo-majorization curve. A graphical description of this proof is shown in Figure 5.1.

As such, the curve in Eq. (5.35) is never below that in Eq. (5.34). This gives us:

$$\tilde{V}_x(\rho) \geq \tilde{V}_x\left(\rho'\right) \geq p\tilde{V}_x(\sigma),$$

where the first inequality holds as, by definition, $\rho$ thermo-majorizes $\rho'$. In particular, we have:

$$p^* \leq \min_{x\in\mathcal{L}(\sigma)} \frac{\tilde{V}_x(\rho)}{\tilde{V}_x(\sigma)},$$

as required.

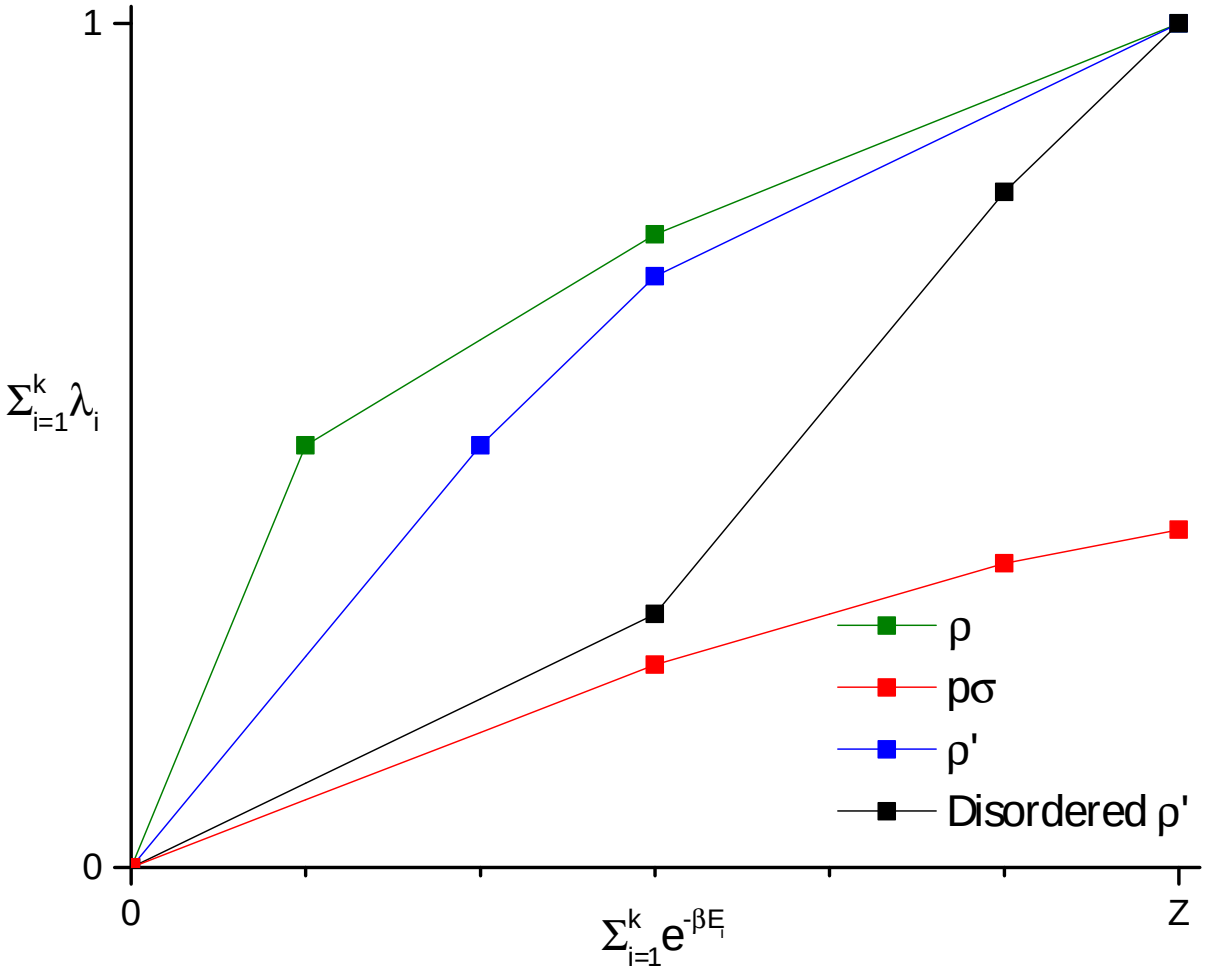



Figure 5.1: *Thermo-majorization diagram for probabilistic thermal operations.* The disordered $\rho'$, Eq. (5.34), has the same $\beta$-ordering as $p\sigma$, Eq. (5.33). By Eq. (5.32), its curve is never below that of $p\sigma$. By construction, the curve of disordered $\rho'$ is always below that of the properly ordered $\rho'$, which, by definition, is itself thermo-majorized by $\rho$.

When $\sigma$ is block-diagonal in the energy eigenbasis, a protocol that saturates the bound is:

$$\begin{aligned}\rho &\xrightarrow{TO} \rho_\sigma, \\ &\xrightarrow{TO} \rho'_\sigma = p^*\sigma + (1-p^*)\,X,\end{aligned}$$

where $\rho_\sigma$ was defined in Eq. (4.27) and is thermo-majorized by $\rho$. As $\rho_\sigma$ and $\sigma$ have the same $\beta$-ordering and:

$$\frac{\tilde{V}_x(\rho)}{\tilde{V}_x(\sigma)} = \frac{\tilde{V}_x(\rho_\sigma)}{\tilde{V}_x(\sigma)}, \quad \forall x \in \mathcal{L}(\sigma),$$

applying the same construction used in Theorem 18 gives a strategy to produce $\rho'_\sigma$ that achieves:

$$p^* = \min_{x\in\mathcal{L}(\sigma)} \frac{\tilde{V}_x(\rho)}{\tilde{V}_x(\sigma)}.$$

□

For block-diagonal $\sigma$, after obtaining $\rho'$ through thermal operations we may apply the measurement defined by Eq. (5.9) to extract our target state with probability $p^*$. This can be done through a process that uses: an ancilla qubit system, $Q$, that starts and ends in the state $|0\rangle$ and has associated Hamiltonian $H_Q = \mathbb{I}_2$, a unitary that correlates the system with the ancilla and a projective measurement on the ancilla qubit. As the measurement operators are diagonal in the energy eigenbasis, we will find that the unitary is energy conserving and within the set of thermal operations. Hence, the only cost we have to pay is to erase the record of the measurement outcome.

The unitary that we shall use is given by:

$$U_{\mathrm{SQ}} = \begin{pmatrix} \sqrt{M} & \sqrt{\mathbb{I}-M} \\ \sqrt{\mathbb{I}-M} & -\sqrt{M} \end{pmatrix}, \tag{5.36}$$

where $M$ is defined as per Eq. (5.9). Note that $U_{\mathrm{SQ}} = U_{\mathrm{SQ}}^\dagger$. Its effect on the initial joint state is:

$$\begin{aligned} U_{\mathrm{SQ}}(\rho' \otimes |0\rangle\langle 0|)U_{\mathrm{SQ}}^\dagger &= \begin{pmatrix} \sqrt{M} & \sqrt{\mathbb{I}-M} \\ \sqrt{\mathbb{I}-M} & -\sqrt{M} \end{pmatrix} \begin{pmatrix} \rho' & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{M} & \sqrt{\mathbb{I}-M} \\ \sqrt{\mathbb{I}-M} & -\sqrt{M} \end{pmatrix}, \\ &= \begin{pmatrix} \sqrt{M}\rho'\sqrt{M} & \sqrt{M}\rho'\sqrt{\mathbb{I}-M} \\ \sqrt{\mathbb{I}-M}\rho'\sqrt{M} & \sqrt{\mathbb{I}-M}\rho'\sqrt{\mathbb{I}-M} \end{pmatrix}, \\ &= \begin{pmatrix} p^*\sigma & \sqrt{M}\rho'\sqrt{\mathbb{I}-M} \\ \sqrt{\mathbb{I}-M}\rho'\sqrt{M} & (1-p^*)X \end{pmatrix}. \end{aligned}$$

If we now measure the ancilla in the computational basis, the joint state will collapse to $\sigma \otimes |0\rangle\langle 0|$ when the 0 outcome is observed. This happens with probability $p^*$. If the 1 outcome is observed, the joint state collapses to $X \otimes |1\rangle\langle 1|$ and this happens with probability $1 - p^*$. In addition, if the 1 outcome is observed, we can then apply a Pauli $Z$ to the ancilla qubit to return it to its initial state.

To see that $U_{SQ}$ commutes with the total Hamiltonian and belongs to the class of thermal operations, first note that the total Hamiltonian is given by:

$$H_{SQ} = H_S \otimes \mathbb{I}_2 + \mathbb{I}_n \otimes \mathbb{I}_2. \tag{5.37}$$

The unitary trivially commutes with the second term so focusing on the first term, and noting that $M$ and $H_S$ are both diagonal matrices so commute, it is easy to check that:

$$\begin{aligned}
[U_{\mathrm{SQ}}, H_S \otimes \mathbb{I}_2] = &\begin{pmatrix} \sqrt{M} & \sqrt{\mathbb{I} - M} \\ \sqrt{\mathbb{I} - M} & -\sqrt{M} \end{pmatrix} \begin{pmatrix} H_{\mathrm{S}} & 0 \\ 0 & H_{\mathrm{S}} \end{pmatrix} \\
&- \begin{pmatrix} H_{\mathrm{S}} & 0 \\ 0 & H_{\mathrm{S}} \end{pmatrix} \begin{pmatrix} \sqrt{M} & \sqrt{\mathbb{I} - M} \\ \sqrt{\mathbb{I} - M} & -\sqrt{M} \end{pmatrix}, \\
&= 0.
\end{aligned}$$

Hence, $[U_{SQ}, H_{SQ}] = 0$.

Observe that this reasoning can be generalized to measurements with $k$ outcomes [131]. Provided the measurement operators commute with $H_S$, the measurement can be performed using a $k$-level ancilla system with trivial Hamiltonian and a joint energy conserving unitary. Such a measurement can be performed for free up to having to spend work to erase the record of the measurement outcome. On the other hand, channels that are not composed of thermal operations (including some measurements characterized by non-diagonal operators) can be seen as a resource [130].

### 5.3.2 Quantifying the work of transition

In thermal operations, the horizontal distance between a state's thermo-majorization curve and the $y$-axis is again a monotone for each value of $y \in [0, 1]$. We denote these by $\tilde{L}_y$ and, as before, they never decrease under thermal operations. In particular, for block-diagonal $\rho$, we

have:

$$\begin{aligned}\tilde{L}_{y_k}(\rho) &= \sum_{i=1}^{k} e^{-\beta E_i^{(\rho)}}, \quad \text{for } y_k = \sum_{i=1}^{k} \eta_i^{(\rho)}, \quad 1 \leq k < \operatorname{rank}(\rho), \\ \tilde{L}_1(\rho) &= \sum_{i=1}^{\operatorname{rank}(\rho)} e^{-\beta E_i^{(\rho)}},\end{aligned} \tag{5.38}$$

where all sums have been properly $\beta$-ordered.

Similarly to Lemma 13, we find:

**Lemma 15.** *Given two states, $\rho$ and $\sigma$, where $\sigma$ is block-diagonal in the energy eigenbasis, under thermal operations:*

$$e^{-\beta W_{\rho\to\sigma}} = \max_{y\in\mathcal{D}(\sigma)} \frac{\tilde{L}_y(\rho)}{\tilde{L}_y(\sigma)}. \tag{5.39}$$

The proof is near identical to that given for noisy operations and so we omit it here. If $\sigma$ is not block-diagonal, the right hand side of Eq. (5.39) lower bounds $e^{-\beta W_{\rho\to\sigma}}$ (to see this, recall that decohering is a thermal operation and hence $W_{\rho\to\sigma} \leq W_{\rho\to\sigma_D}$). Note that the work of transition can also be calculated in terms of a linear program [148].

For a changing Hamiltonian modeled using Eq. (4.33), $\tilde{L}_y(\rho) = \tilde{L}_y(\rho\otimes|0\rangle\langle 0|)$ for $0 \leq y \leq 1$ (and similarly for $\sigma$). Hence, changing Hamiltonian does not affect the above calculation.

### Bounds on the transition probability

We can also prove a result analogous to Lemma 14 for thermal operations:

**Lemma 16.** *Given two states, $\rho$ and $\sigma$, where $\sigma$ is block-diagonal in the energy eigenbasis and $W_{\rho\to\sigma} \leq 0$, then under thermal operations:*

$$e^{\beta W_{\rho\to\sigma}} \leq p^* \leq e^{-\beta W_{\sigma\to\rho_D}}. \tag{5.40}$$

*If $W_{\rho\to\sigma} \geq 0$, $p^* = 1$ and the transformation from $\rho$ to $\sigma$ can be done deterministically, potentially extracting a finite amount of work.*

*Proof.* The proof of the lower bound is best illustrated using thermo-majorization curves, and is given in Figure 5.2.

For the upper bound, assume that $W_{\sigma\to\rho_D} \geq 0$ so that the the bound is not trivial. Let:

$$\begin{aligned}M &= \max_{i\in\{1,\ldots,n\}} \zeta_i e^{\beta E_i}, \\ N &= \max_{i\in\{1,\ldots,n\}} \eta_i e^{\beta E_i},\end{aligned}$$

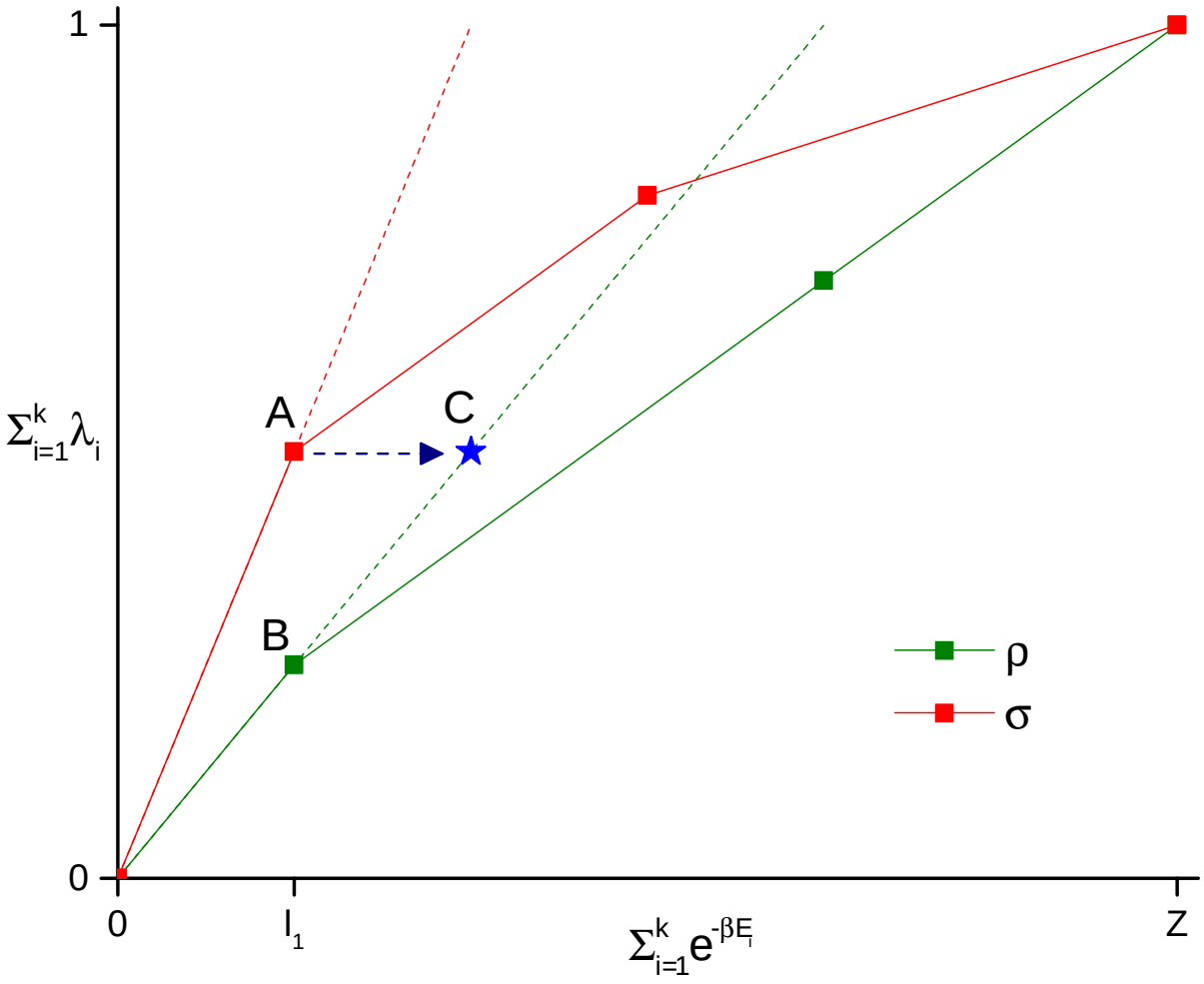



Figure 5.2: *Lower bounding the transition probability.* To prove the lower bound in Lemma 16, let $l_1$ be a point such that $p^* = \frac{\tilde{V}_{l_1}(\rho)}{\tilde{V}_{l_1}(\sigma)}$. Diagrammatically, $p^*$ is calculated by taking the ratio of the heights of the thermo-majorization curves at $B$ and $A$. If the curve of $\sigma$ is stretched by a factor of $\frac{1}{p^*}$ relative to the $x$-axis, then this will take the point $A$ to the point $C$. As illustrated in Figure 4.7, $W_{\rho\to\sigma}$ is such that stretching the thermo-majorization curve of $\sigma$ by a factor of $e^{-\beta W_{\rho\to\sigma}}$ places the thermo-majorization curve of $\sigma$ just to the right of that of $\rho$. Hence, we have $e^{\beta W_{\rho\to\sigma}} \leq p^*$, as required.

where $\zeta_i$ and $\eta_i$ are the eigenvalues associated with energy level $E_i$ for $\sigma$ and $\rho_D$ respectively. Using Eqs. (5.31) and (5.39), we have:

$$p^* \leq \frac{N}{M} \leq e^{-\beta W_{\sigma \to \rho_D}},$$

and we obtain the bound. □

## 5.4 Summary

In this chapter we defined and calculated the probability of making a transition between two states under thermal operations. This can be done using a finite set of the monotones, $\tilde{V}_x$, which are defined by the height of the thermo-majorization curves. A similar calculation involving a finite number of the monotones $\tilde{L}_y$, defined by the distance between the thermo-majorization curve and the $y$-axis, yields the work of transition. Finally, we saw that the work of transition between the two states, and vice versa, can be used to bound the probability of making the transition.

Currently, as previously highlighted in Section 4.3.2, little is known about the case when the final state is not block-diagonal in the energy eigenbasis. In such a situation, our results provide upper bounds that are not necessarily achievable. Determining the achievable values for both $p^*$ and $W_{\rho \to \sigma}$ for this case is expected to be difficult, as we do not yet know when a transition is possible for non-probabilistic transformations.

The analysis in this chapter has focused on noisy and thermal operations in the absence of a catalyst. Having access to catalysts has the potential to achieve higher values of $p$ than that defined by $p^*$ and it would be interesting to find an expression or bound for the maximum $p$ in the process:

$$\rho \xrightarrow{\text{CTO}} \rho' = p\sigma + (1-p)\, X. \tag{5.41}$$

Note that a bound can be obtained from any non-increasing monotone of catalytic thermal operations, $M$ say, that satisfies:

$$M\left(p\sigma + (1-p)\, X\right) \geq pM\left(\sigma\right). \tag{5.42}$$

Bounding the maximum transition probability under catalytic thermal operations is made more difficult by the fact that the generalized free energies found in [24] are not all concave.

In maximizing the value of $p$ in Eq. (5.1) to obtain $p^*$, we have attempted to maximize the fraction of $\sigma$ present in a state obtainable from $\rho$. With access to a single two outcome

measurement, $\sigma$ can also be obtained from $\rho$ with probability $p^*$. There are other measures that one could quantify in attempting to obtain a state that behaves like $\sigma$. For example, one could consider the fidelity between $\sigma$ and a state reachable from $\rho$:

$$F_{TO}(\rho,\sigma) \equiv \max_{\tilde{\rho}} \left\{ F(\tilde{\rho},\sigma) : \rho \xrightarrow{TO} \tilde{\rho} \right\}, \tag{5.43}$$

where $F(\tilde{\rho},\sigma) = \mathrm{Tr}\left[\sqrt{\sqrt{\tilde{\rho}}\sigma\sqrt{\tilde{\rho}}}\right]$ is the fidelity between the two states. Investigating this problem is an open question, but note that for diagonal $\sigma$ we have $F_{TO}(\rho,\sigma) \geq F(\rho',\sigma) \geq \sqrt{p^*}$.

Another avenue of research is to generalize our result to the case where one is interested in not only maximizing the probability of obtaining a single state but instead finding the probability simplex of going to an ensemble of many states. Again, the fact that the monotones used in thermodynamics are not in general concave, means that a straight application of the techniques used in entanglement theory [100] cannot be immediately applied.

Finally, by supplying some initial work or demanding that extra work be extracted, the value of $p^*$ achieved can be raised or lowered. What is the tradeoff between $p^*$ and this work, $W$? As an example, the solution for qubit systems in the framework of noisy operations is given in Appendix C while the more general question has been considered in [149].

As seen in this chapter, the structure and geometry of thermo-majorization curves provide a useful tool for studying nano-scale thermodynamics. In the next chapter, we shall use them further to analyze what is possible under a restricted set of thermodynamical manipulations.

# Chapter 6

# Towards Experimentally Friendly Thermal Operations

## 6.1 Coarse operations

The constraints on nano-scale thermodynamics captured by thermo-majorization apply even if one is allowed to manipulate the microscopic degrees of freedom of both the system and the heat bath. Furthermore, at least when considering systems without coherence, it is known that such precise control enables one to perform those transformations that thermo-majorization does not rule out. However, being able to address the individual micro-states of both system and heat bath in such a manner and apply the full range of unitaries allowed under thermal operations is beyond the current reach of experiment.

Achieving a macroscopic state transition that is allowed by the traditional laws of thermodynamics does not require such fine-grained command. Is it truly necessary when considering small systems? In this chapter, we shall introduce two simple operations, much more applicable to experimental test. Surprisingly, if these operations are complemented by the ability to similarly manipulate a single thermal qubit (as opposed to the entirety of the heat bath), we will show that they that allow one to perform all transitions between diagonal states that can be implemented using thermal operations.[1]

Due to the level of control they require, we shall refer to this set of simple operations as

[1] In fact, if we supplement these operations with the ability to perform any energy conserving unitary on the system (those $U$ such that $[H_S, U] = 0$), then these will be enough to convert any $\rho$ into any block-diagonal $\sigma$ when the transformation is allowed by thermo-majorization.

*coarse operations*. The ability to perform any energy conserving unitary across system and bath is replaced with the primitives of *Partial Level Thermalizations* and *Level Transformations*.

Partial Level Thermalizations probabilistically thermalize the state of the system over a subset of energy levels. Level Transformations on the other hand, consist of raising and lowering the energy levels of the Hamiltonian and the work cost of implementing them is related to the change in energy of populated levels. A similar set of operations was considered in [2] where it was shown that Level Transformations combined with the ability to fully thermalize the system is sufficient to distill the optimal amount of work from a state. By allowing slightly more control over thermalization and the use of an ancilla thermal qubit, we will see that much more can be achieved.

More fully, in contrast to thermal operations, under coarse operations the following processes can be applied:

1. A single 2-level system with known Hamiltonian, in the Gibbs state of that Hamiltonian at temperature $T$, can be appended as an ancilla.

2. Partial Level Thermalization (PLT) can be implemented over any subset of energy levels.

3. Level Transformations (LT), can be performed provided any associated work cost is accounted for.

4. The ancilla system may be discarded.

Operations 1 and 4 are reminiscent of Operations 1 and 3 introduced in Section 4.2.2 for thermal operations. Note however, that while in thermal operations an ancilla system with arbitrary Hamiltonian can be used, under coarse operations, we are restricted to one, 2-level Hamiltonian in particular. Furthermore, it does not matter what the Hamiltonian of this qubit system is, provided it is known.

In the remainder of this section, we shall formally define Operations 2 and 3 of coarse operations and introduce some useful protocols that can be formed from combinations of the two.

### 6.1.1 Partial Level Thermalizations

Partial Level Thermalizations adjust the occupation probabilities of the system's state for some subset of the system's energy levels. With some probability, $\lambda$, a PLT thermalizes the system

over, and with respect to, this subset of levels, while with probability $1-\lambda$, it leaves them unchanged. More formally:

**Definition 27** (Partial Level Thermalization)**.** *Given an n-level system with Hamiltonian* $H_S = \sum_{i=1}^{n} E_i|i\rangle\langle i|$*, a Partial Level Thermalization is parametrized by* $\lambda \in [0,1]$ *and acts on some subset of the system's energy levels. Denote this subset of energy levels by* $\mathcal{P}$ *and the Partial Level Thermalization by* $PLT_{\mathcal{P}}(\lambda)$*.*

*The action of* $PLT_{\mathcal{P}}(\lambda)$ *on* $\rho = \sum_{i=1}^{n} \eta_i|i\rangle\langle i|$*, is defined by:*

$$\rho \overset{PLT_{\mathcal{P}}(\lambda)}{\longrightarrow} \rho', \tag{6.1}$$

*where* $\rho' = \sum_{i=1}^{n} \eta_i'|i\rangle\langle i|$ *and the* $\eta_i'$ *are such that, for* $i \in \mathcal{P}$*:*

$$\eta_i' = (1-\lambda)\,\eta_i + \frac{\lambda e^{-\beta E_i}}{\sum_{i\in\mathcal{P}} e^{-\beta E_i}} \sum_{i\in\mathcal{P}} \eta_i, \tag{6.2}$$

*and* $\eta_i' = \eta_i$ *otherwise.*

The action of a Partial Level Thermalization is illustrated in terms of thermo-majorization curves in Figure 6.1.

Such an operation preserves the $\beta$-ordering of the levels in $\mathcal{P}$ as, for $i,j \in \mathcal{P}$, if $\eta_i e^{\beta E_i} \geq \eta_j e^{\beta E_j}$, then $\eta_i' e^{\beta E_i} \geq \eta_j' e^{\beta E_j}$. In particular, if for some $d$, $\mathcal{P} = \left\{i^{(\rho)}\right\}_{i=k}^{k+d-1}$ (with the superscript $(\rho)$ denoting that the energy levels are $\beta$-ordered with respect to $\rho$), then $\rho'$ will have the same $\beta$-ordering as $\rho$.

### 6.1.2 Level Transformations

In contrast to Partial Level Thermalizations, Level Transformations adjust the energy levels of a system's Hamiltonian while preserving the occupation probabilities of a state. Furthermore, they may cost work to implement. The action of a Level Transformation is captured in the following definition:

**Definition 28** (Level Transformation)**.** *Given an n-level system with Hamiltonian* $H_S = \sum_{i=1}^{n} E_i|i\rangle\langle i|$ *in the state* $\rho = \sum_{i=1}^{n} \eta_i|i\rangle\langle i|$*, a Level Transformation is parametrized by a set of real numbers,* $\mathcal{E} = \{h_i\}_{i=1}^{n}$*, and denoted by* $LT_{\mathcal{E}}$*.*

*The action of* $LT_{\mathcal{E}}$ *on* $(\rho, H_S)$ *is:*

$$(\rho, H_S) \overset{LT_{\mathcal{E}}}{\longrightarrow} \left(\rho, H_S'\right), \tag{6.3}$$

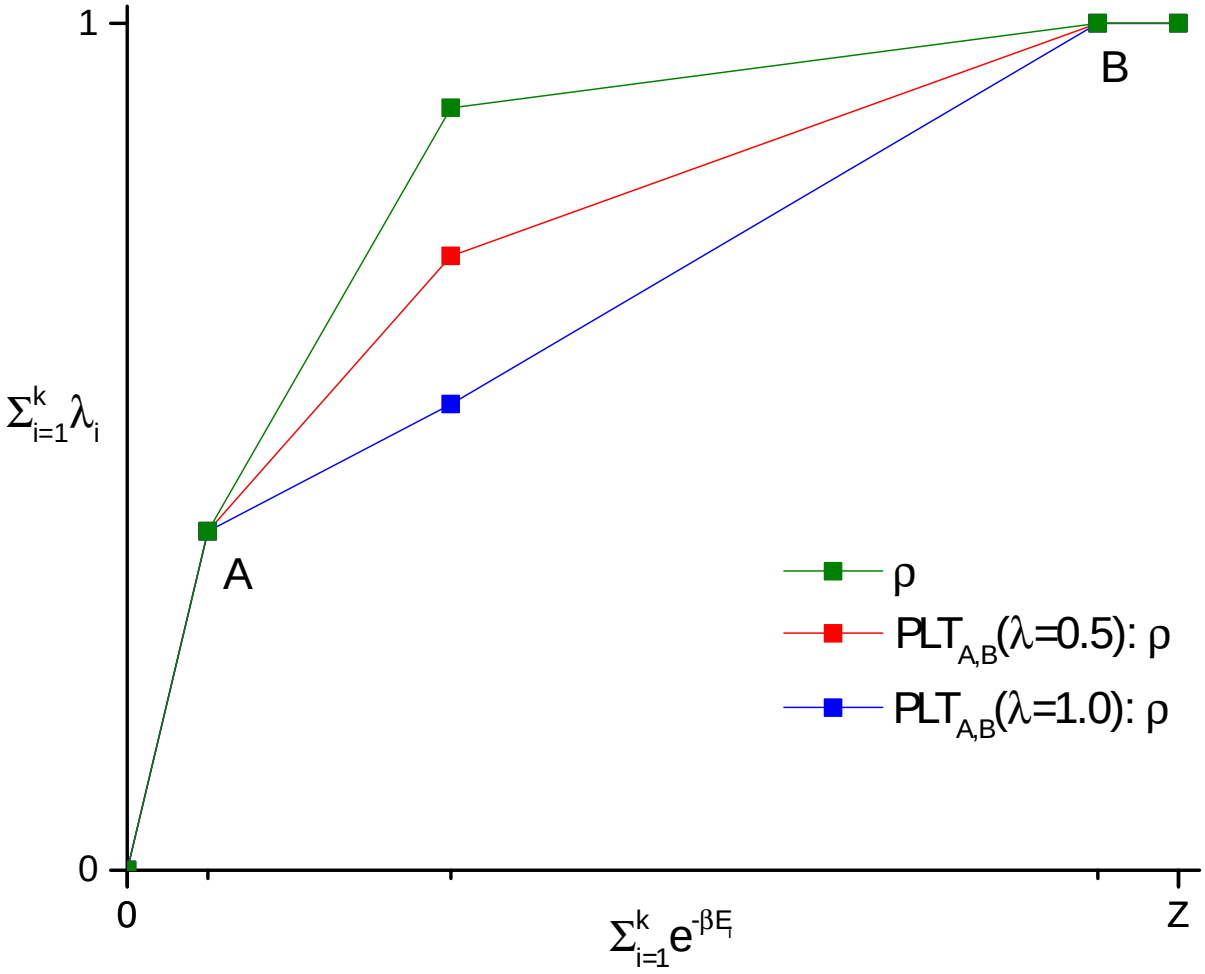



Figure 6.1: *Partial Level Thermalization.* The action of a PLT applied to a state, $\rho$, over the two energy levels between points $A$ and $B$ for various choices of $\lambda$. Note, that as we apply the PLT only to adjacent energy levels with respect to the $\beta$-ordering of $\rho$, the final states maintain this $\beta$-ordering.

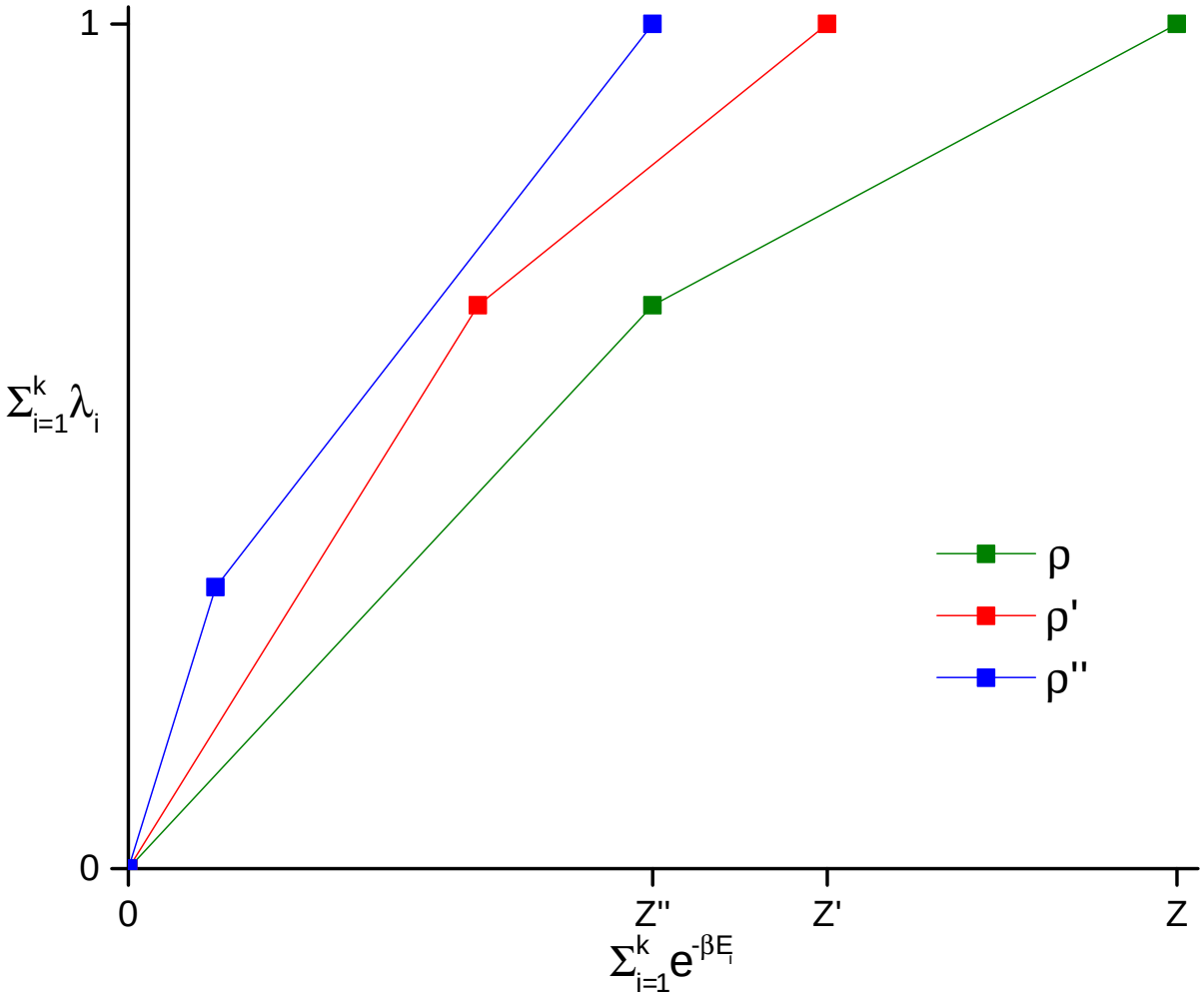



Figure 6.2: *Level Transformation.* The action of LTs applied to a state $\rho$. Note that LTs leave the occupation probabilities unchanged but may alter the $\beta$-ordering of a state.

*where:*

$$H'_S = \sum_{i=1}^{n} (E_i + h_i) \, |i\rangle\langle i|. \tag{6.4}$$

*The single-shot, worst-case, work cost/yield of $LT_{\mathcal{E}}$ is defined to be:*

$$W_{LT_{\mathcal{E}}} = -\max_{i:\eta_i>0} h_i. \tag{6.5}$$

*If $W_{LT_{\mathcal{E}}}$ is negative, work must be added for the transformation to happen deterministically while if it is positive, it may be possible to extract some work.*

The action of a Level Transformation is illustrated in terms of thermo-majorization curves in Figure 6.2.

### 6.1.3 Partial Isothermal Reversible Processes and Points Flows

We can combine sequences of Level Transformations and Partial Level Thermalizations in such a way to form a useful protocol, termed a *Partial Isothermal Reversible Process* (PITR) as they are similar in construction to the Isothermal Reversible Processes considered in [2] but require Partial Level Thermalizations rather than full thermalizations. In terms of thermo-majorization curves, Partial Isothermal Reversible Processes will enable us to move non-elbow points along the segments on which they exist, without changing the shape and structure of the rest of the curve. More formally:

**Definition 29** (Partial Isothermal Reversible Process)**.** *Given an $n$-level system with Hamiltonian $H_S = \sum_{i=1}^{n} E_i|i\rangle\langle i|$, a Partial Isothermal Reversible Process is parametrized by a positive constant, $\kappa$, and acts on some pair of the system's energy levels, indexed by $j$ and $k$. Denote the Partial Isothermal Reversible Process by $PITR_{j,k}(\kappa)$.*

*The action of $PITR_{j,k}(\kappa)$ on $(\rho, H_S)$, where $\rho = \sum_{i=1}^{n} \eta_i|i\rangle\langle i|$, is defined by:*

$$(\rho, H_S) \overset{PITR_{j,k}(\kappa)}{\longrightarrow} \left(\rho', H'_S\right), \tag{6.6}$$

*where $\rho' = \sum_{i=1}^{n} \eta'_i$ and $H'_S = \sum_{i=1}^{n} E_i|i\rangle\langle i|$. Defining $\tilde{\eta}_j = \frac{e^{-\beta E_j}}{e^{-\beta E_j}+e^{-\beta E_k}}(\eta_j + \eta_k)$ and $\tilde{\eta}_k = \frac{e^{-\beta E_k}}{e^{-\beta E_j}+e^{-\beta E_k}}(\eta_j + \eta_k)$, the components of $(\rho', H'_S)$ in terms of $\kappa$ are then:*

$$\begin{aligned}
\eta'_j &= \tilde{\eta}_j e^{-\beta\kappa}, \\
\eta'_k &= \tilde{\eta}_k + \left(1 - e^{-\beta\kappa}\right)\tilde{\eta}_j, \\
E'_j &= E_j + \kappa, \\
E'_k &= -\frac{1}{\beta}\ln\left[e^{-\beta E_k} + e^{-\beta E_j}\left(1 - e^{-\beta\kappa}\right)\right],
\end{aligned} \tag{6.7}$$

*with $\eta'_i = \eta_i$ and $E_i = E'_i$ for $i \notin \{j, k\}$.*

A Partial Isothermal Reversible Process is illustrated in terms of thermo-majorization diagrams in Figure 6.3. Note that in such a process, $Z_S = Z'_S$ and that, for all $\kappa$:

$$\eta'_j e^{\beta E'_j} = C = \eta'_k e^{\beta E'_k}, \tag{6.8}$$

where $C$ is some constant.

That the action defined in Definition 29 can be performed using a protocol consisting only of PLTs and LTs is shown in the proof of the following lemma. Furthermore, such a protocol costs no work with probability close to one:

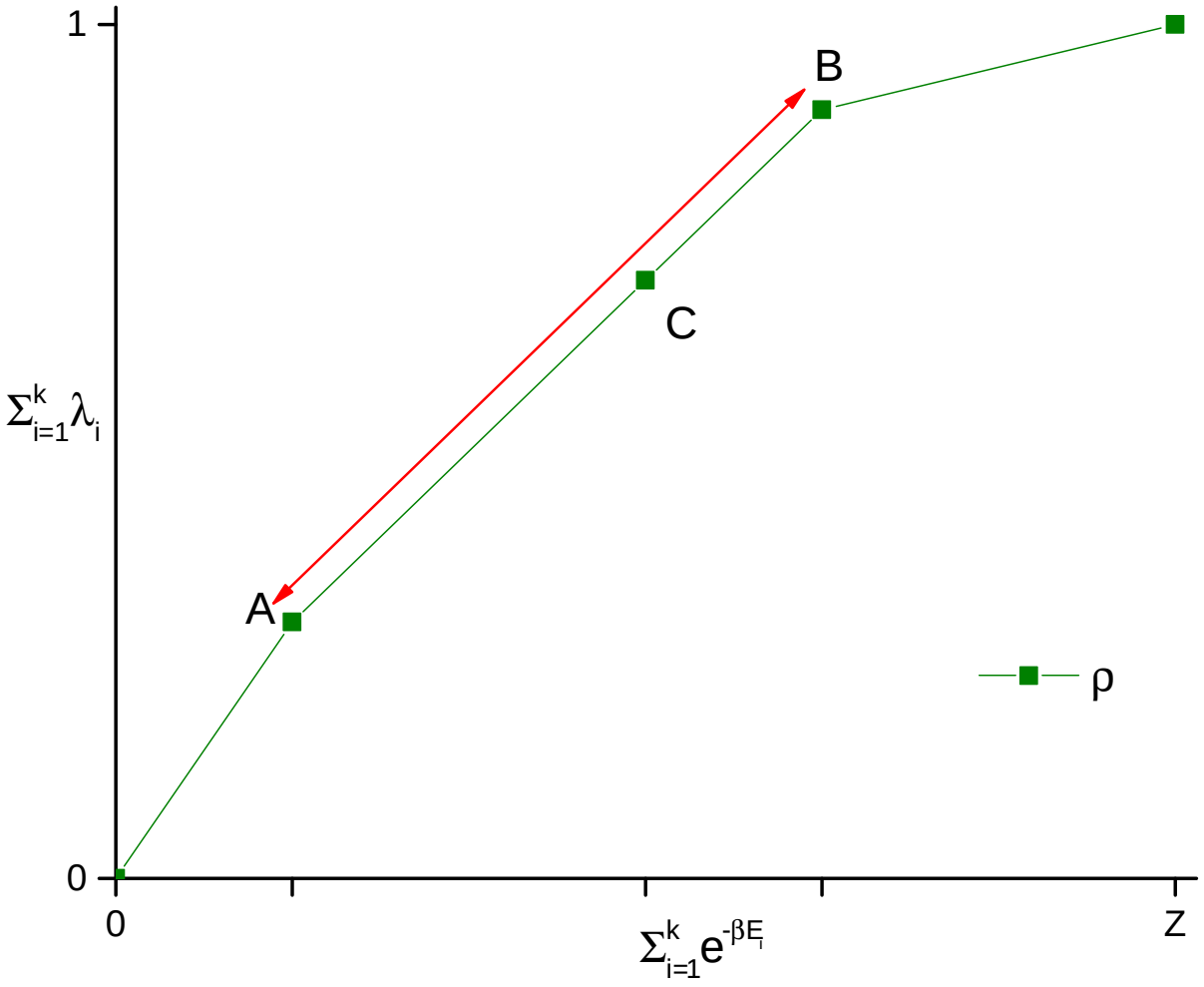



Figure 6.3: *Partial Isothermal Reversible Process.* The action of a PITR applied to the system $(\rho, H)$ on the two energy levels between points $A$ and $B$. By choosing the value of $\kappa$, the point $C$ can be moved such that it lies anywhere on the line-segment between $A$ and $B$ without changing the overall shape of the thermo-majorization curve.

**Lemma 17.** *The operation $PITR_{j,k}(\kappa)$ can be performed using Partial Level Thermalizations and Level Transformations. In addition, such a protocol can have work cost arbitrarily close to zero with probability arbitrarily close to one.*

*Proof.* To show this, we define a $t$-step procedure that implements $\mathrm{PITR}_{j,k}(\kappa)$ with each step consisting of a Level Transformation and a Partial Level Thermalization. Let $(\rho, H_S)$ and $(\rho', H'_S)$ be defined as per Definition 29. Without loss of generality, assume that:

$$\eta_j e^{\beta E_j} = C = \eta_k e^{\beta E_k}.$$

for some constant $C$ (if not, we can always perform the PLT, $\mathrm{PLT}_{\{j,k\}}(\lambda = 1)$ to make it so and, as we shall see in Lemma 18, this is a thermal operation so costs no work). Hence, in the terminology of Definition 29, $\eta_j = \tilde{\eta}_j$ and $\eta_k = \tilde{\eta}_k$.

Define $\epsilon = \frac{E'_j - E_j}{t}$. Let the Hamiltonian after step $r$ be $H_S^{(r)} = \sum_{i=1}^n E_i^{(r)} |i\rangle\langle i|$ and the state of the system be $\rho^{(r)} = \sum_{i=1}^n \eta^{(r)} |i\rangle\langle i|$. In step $r$, we perform the level transformation such that:

$$\begin{aligned} E_j^{(r)} &= E_j + r\epsilon, \\ E_k^{(r)} &= -\frac{1}{\beta} \ln \left[ e^{-\beta E_k} + e^{-\beta E_j} \left( 1 - e^{-\beta r \epsilon} \right) \right], \end{aligned}$$

and fully thermalize over energy levels $j$ and $k$ so that:

$$\begin{aligned} \eta_j^{(r)} &= \eta_j e^{-\beta r \epsilon}, \\ \eta_k^{(r)} &= \eta_k + \left( 1 - e^{-\beta r \epsilon} \right) \eta_j. \end{aligned}$$

All other energy levels and occupation probabilities remain unchanged. It can readily be verified that $Z_S^{(r)} = Z_S$ and that:

$$\eta_j^{(r)} e^{\beta E_j^{(r)}} = C = \eta_k^{(r)} e^{\beta E_k^{(r)}}.$$

Hence, this protocol produces the desired $(\rho', H'_S)$ after $t$ steps.

Such a protocol alters the state and Hamiltonian of the system but does not change the shape of the system's thermo-majorization curve. Following the proof of [2, Supplementary Lemma 1.] regarding Isothermal Reversible Processes, we shall now show that the work cost of this protocol becomes increasingly peaked around zero as the number of steps taken tends to infinity.

Let $W^{(r)}$ denote the random variable for the work distribution in step $r$ of the PITR. The work distribution for the whole $t$-step PITR process is then:

$$W^{\text{PITR}} = \sum_{r=1}^{t} W^{(r)}.$$

Using Eqs. (6.7) and (6.8), $W^{(r)}$ is such that with probability:

$$\begin{aligned}
&Ce^{-\beta E_j^{(r)}}, \quad W^{(r)} = \epsilon, \\
&Ce^{-\beta E_k^{(r)}}, \quad W^{(r)} = -\frac{1}{\beta} \ln \left[ e^{-\beta E_k^{(r)}} + e^{-\beta E_j^{(r)}} \left(1 - e^{-\beta\epsilon}\right) \right] - E_k^{(r)}, \\
&\text{otherwise}, \quad W^{(r)} = 0.
\end{aligned}$$

Now, for large $t$ (small $\epsilon$) this becomes such that with probability:

$$\begin{aligned}
&Ce^{-\beta E_j^{(r)}}, \quad W^{(r)} = \epsilon, \\
&Ce^{-\beta E_k^{(r)}}, \quad W^{(r)} = -\epsilon e^{\beta\left(E_k^{(r)} - E_j^{(r)}\right)} + O(\epsilon^2), \\
&\text{otherwise}, \quad W^{(r)} = 0.
\end{aligned}$$

Hence as $t \to \infty$, $\left\langle W^{(r)} \right\rangle \to 0$, for all $r$. As $t = \frac{E_j^{(t)} - E_j^{(0)}}{\epsilon} = \frac{E'_j - E_j}{\epsilon}$, we have that:

$$\left\langle W^{\text{PITR}} \right\rangle = \sum_{r=1}^{t} \left\langle W^{(r)} \right\rangle \to 0, \quad \text{as } t \to \infty.$$

Now consider the variance of $W^{\text{PITR}}$. For large $t$, hence small $\epsilon$:

$$\begin{aligned}
\left\langle {W^{(r)}}^2 \right\rangle &= Ce^{-\beta E_j^{(r)}} \epsilon^2 + Ce^{-\beta E_k^{(r)}} \epsilon^2 e^{2\beta\left(E_k^{(r)} - E_j^{(r)}\right)} + O(\epsilon^4) \\
&= Ce^{-\beta E_j^{(r)}} \epsilon^2 \left(1 + e^{\beta\left(E_k^{(r)} - E_j^{(r)}\right)}\right) + O(\epsilon^4) \\
&\to 0 \quad \text{as } \epsilon \to 0.
\end{aligned}$$

Hence, $\text{Var}\left(W^{(r)}\right) \to 0$ as $t \to \infty$. As the $W^{(r)}$ are independent:

$$\begin{aligned}
\text{Var}\left(W^{\text{PITR}}\right) &= \sum_{r=1}^{t} \text{Var}\left(W^{(r)}\right) \\
&\to 0 \quad \text{as } \epsilon \to 0.
\end{aligned}$$

Note that this analysis extends to the case where $E'_j \to \infty$. If we parametrize $E'_j$ in terms of the number of steps taken in the PITR protocol so that $E'_j = \ln t$, then in the limit $t \to \infty$, $E'_j \to \infty$, $\left\langle W^{\text{PITR}} \right\rangle \to 0$ and $\text{Var}\left(W^{\text{PITR}}\right) \to 0$.

Now, Chebyshev's inequality gives us that:

$$P\left(\left|W^{\mathrm{PITR}}\right| \geq K\sqrt{\mathrm{Var}\left(W^{\mathrm{PITR}}\right)}\right) \leq \frac{1}{K^2},$$

so by taking $t$ and $K$ to be large, we obtain that the work distribution for the PITR becomes increasingly peaked around 0. □

Hence, Partial Isothermal Reversible Processes can be used to move non-elbow points along straight line-segments of a thermo-majorization curve, using coarse operations and without expending any work with high probability. As such, in a slight abuse of terminology, we shall say that they allow us to move points in this fashion with essentially no work cost. By combining two PITRs, it is possible to commute non-elbow points with elbows, meaning that non-elbows can be moved to any point of the thermo-majorization curve for free. We term this operation *Exact Points Flow*:

**Definition 30** (Exact Points Flow (EPF))**.** *Given an n-level system with Hamiltonian* $H_S = \sum_{i=1}^{n} E_i|i\rangle\langle i|$ *and state* $\rho = \sum_{i=1}^{n} \eta_i|i\rangle\langle i|$ *such that:*

$$\eta_j e^{\beta E_j} = \eta_k e^{\beta E_k}, \tag{6.9}$$

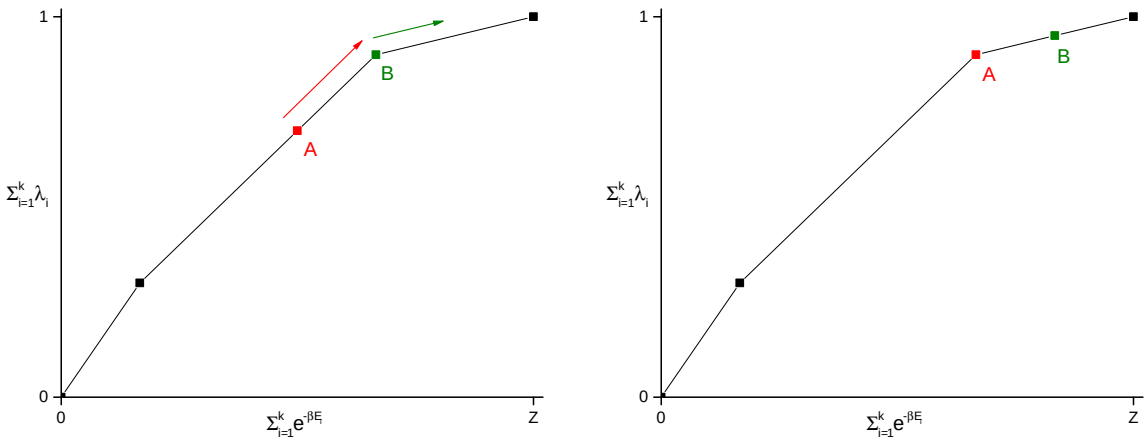
*for some* $j, k$ *(i.e. there is a non-elbow point on the thermo-majorization curve), an Exact Points Flow moves this non-elbow to another part of the thermo-majorization curve whilst keeping the shape of the curve fixed.*

Exact Points Flow is illustrated in Figure 6.4. To implement it, we first apply a PITR that sends $E_j \to \infty$. This lowers the energy of $E_k$ and does not alter the shape of the thermo-majorization curve. Next, to move the non-elbow to another part of the curve, we apply another PITR to $j$ and a third level labeled by $l$, bringing the energy level associated with $j$ back down from infinity to some $E'_j$. This leaves us with a system, $(\rho', H'_S)$, with thermo-majorization curve identical to that of $(\rho, H_S)$ and such that:

$$\eta'_j e^{\beta E'_j} = \eta'_l e^{\beta E'_l}, \tag{6.10}$$

(i.e. the elbow defined in Eq. (6.9) has moved to another part of the curve).

Exact Points Flow requires that an energy level is raised to infinity during a Partial Isothermal Reversible Process. If it is not possible, or undesirable, to raise an energy level to infinity, a similar effect to an Exact Points Flow can be achieved while altering the shape of the thermo-majorization curve slightly in a process we call *Approximate Points Flow*.




(a) Initial state. (b) Final state.

Figure 6.4: *Exact Points Flow.* EPF allows us to move a non-elbow point through the elbow. First, we perform a PITR that sends the non-elbow, $A$, towards the elbow, $B$. As the appropriate energy level is raised to infinity during the protocol, $A$ tends towards $B$ until they coincide. Next, a second PITR lowers the energy level from infinity, keeping it in partial thermal equilibrium with respect to another line-segment. This moves $B$, now a non-elbow, to this new line-segment.

**Definition 31** (Approximate Points Flow (APF))**.** *Given an $n$-level system with Hamiltonian $H_S = \sum_{i=1}^{n} E_i|i\rangle\langle i|$ and state $\rho = \sum_{i=1}^{n} \eta_i|i\rangle\langle i|$ such that:*

$$\eta_j e^{\beta E_j} = \eta_k e^{\beta E_k}, \tag{6.11}$$

*for some $j, k$ (i.e. there is a non-elbow point on the thermo-majorization curve), an Approximate Points Flow moves this non-elbow to an adjacent segment of the thermo-majorization curve whilst modifying the shape of the thermo-majorization curve by an arbitrarily small amount and without sending an energy level to infinity.*

Approximate Points Flow is illustrated in Figure 6.5. To implement it, without loss of generality, assume that the set $\{i\}_{i=1}^{n}$ has been $\beta$-ordered, we wish to move the non-elbow point to the right and take $j = k + 1$. To do this, we:

1. Apply a PITR that raises the energy level of $E_j$ to some fixed, large but finite amount. This lowers the energy of $E_k$ and does not alter the shape of the thermo-majorization curve.

2. We now apply:

$$\mathrm{PLT}_{\{j,j+1\}}\left(\lambda = 1\right), \tag{6.12}$$

to the system. This turns the non-elbow associated with $j$ and $k$ into an elbow and the elbow associated with $j$ and $j+1$ into a non-elbow.

3. Using a PITR, we can now move the new non-elbow point without altering the shape of the thermo-majorization curve.

(i.e. the elbow defined in Eq. (6.11) has moved to another part of the curve). By adjusting the height to which $E_j$ is raised in Step 1, we can tune the extent to which the thermo-majorization curve is altered.

## 6.2 Coarse operations as thermal operations

The coarse operations introduced in the previous section, together with the protocols based upon them, will be used to implement transformations that are possible under thermal operations. Before doing this, we first show that coarse operations are in fact a subset of thermal operations and that we are not allowing processes from outside of the resource theory.

### 6.2.1 Partial Level Thermalizations

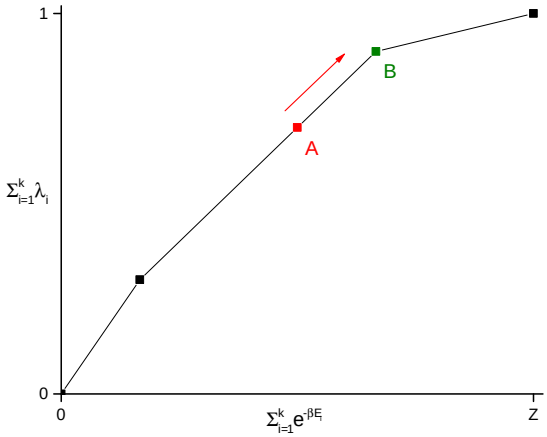
That Partial Level Thermalization form a subset of thermal operations, is captured in the following lemma.

**Lemma 18.** *The map $PLT_{\mathcal{P}}\left(\lambda\right)$ can be implemented using thermal operations.*

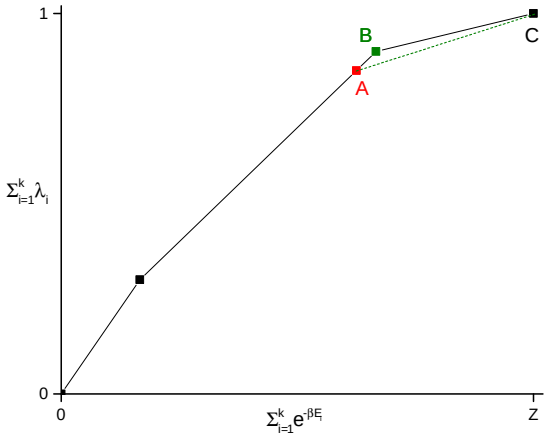
*Proof.* To show this, we give an explicit protocol implementing a Partial Level Thermalization. For simplicity, we assume $\lambda$ is a positive rational of the form $\frac{a}{b}$ (if $\lambda$ is irrational, then $a$ and $b$ should be chosen such that the Partial Level Thermalization is implemented to the desired accuracy). Let the state of the system be $\rho = \sum_{i=1}^{n} \eta_i |i\rangle\langle i|$ and the associated Hamiltonian, $H_S = \sum_{i=1}^{n} E_i |i\rangle\langle i|$. The protocol then runs as follows:

- Step 1. Using Operation 1 of thermal operations:

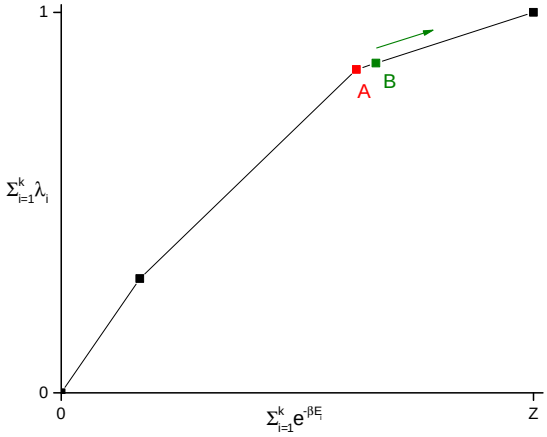
$$\left(\rho, H_S\right) \to \left(\rho \otimes \tau_A \otimes \mathbb{I}_b, H_S + H_A + H_M\right),$$




(a) Initial state.




(b) After first PITR.




(c) Final state.

Figure 6.5: *Approximate Points Flow.* Initially the system is as per Figure (a). Using a PITR, the non-elbow point, $A$, is moved towards the elbow at point $B$. This results in Figure (b). Next, a PLT is applied between points $A$ and $C$, leading to Figure (c). Point $B$ is now a non-elbow and can be moved using a PITR.

where $\tau_A$ is the Gibbs state of a $|\mathcal{P}|$-level system with Hamiltonian:

$$H_A = \sum_{j \in \mathcal{P}} E_j |j\rangle\langle j|,$$

and $\mathbb{I}_b$ is the maximally mixed state of dimension $b$. This is a Gibbs state of a $b$-level system with Hamiltonian $H_M = 0$.

- Step 2. Let $\{|r_S, s_A, t_M\rangle\}$ be the set of orthonormal eigenvectors of $H_S + H_A + H_M$, each with associated energy level $E_r + E_s$. The eigenvalue of $\rho \otimes \tau_A \otimes \mathbb{I}_b$ associated with each energy level is $\frac{1}{bZ_A}\eta_r e^{-\beta E_s}$. Let $U$ be a unitary acting on the global system such that, for $r \in \mathcal{P}$, $\forall s,\, t \in \{1, \ldots, a\}$:

  $$U|r_S, s_A, t_M\rangle = |s_S, r_A, t_M\rangle,$$

  and $U|r_S, s_A, t_M\rangle = |r_S, s_A, t_M\rangle$ otherwise. By construction, $U$ is an energy conserving unitary that commutes with the total Hamiltonian.

- Step 3. Discard the two ancilla systems.

After applying this protocol, the population of energy level $E_j$ for $j \in \mathcal{P}$ is:

$$\begin{aligned} &\eta_j - \frac{a}{bZ_A}\sum_{i \in \mathcal{P}} e^{-\beta E_i}\eta_j + \frac{a}{bZ_A}\sum_{i \in \mathcal{P}} e^{-\beta E_j}\eta_i, \\ &= \left(1 - \frac{a}{b}\right)\eta_j + \frac{a}{b}\frac{e^{-\beta E_j}}{\sum_{i \in \mathcal{P}} e^{-\beta E_i}}\sum_{i \in \mathcal{P}} \eta_i, \end{aligned}$$

and $\eta_j$ otherwise. Comparing this with Eq. (6.2), we see that we have implemented the Partial Level Thermalization as required. □

### 6.2.2 Level Transformations

The work cost of Level Transformations can be modeled within thermal operations using the construction for changing Hamiltonians given in Eq. (4.34).

**Lemma 19.** *The map $LT_{\mathcal{E}}$ can be implemented using thermal operations with work cost at most $W_{LT_{\mathcal{E}}}$.*

*Proof.* Let $H = \sum_{i=1}^n E_i|i\rangle\langle i|$ be the initial Hamiltonian and $H' = \sum_{i=1}^n E'_i|i\rangle\langle i|$ be the final Hamiltonian after the application of $\mathrm{LT}_{\mathcal{E}}$. Let $\mathcal{E} = \{h_i\}_{i=1}^n$ so $E'_i = E_i + h_i$.

Consider modeling this transformation using the switch qubit construction:

$$H_T = H \otimes |0\rangle\langle 0| + H' \otimes |1\rangle\langle 1|.$$

Let $H_W = W|1\rangle\langle 1|$ be the Hamiltonian for the work storage system. The work require to implement the Level Transformation and convert $(\rho, H)$ into $(\rho, H')$ under thermal operations, $W_{H\to H'}$, is then given by the largest value of $W$ such that:

$$(\rho \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|, H_T + H_W) \xrightarrow{\text{TO}} (\rho \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1|, H_T + H_W)\,.$$

To see that $W_{H\to H'} \geq W_{\text{LT}_\mathcal{E}}$, consider the Level Transformation parametrized by $\tilde{\mathcal{E}} = \left\{\tilde{h}_i\right\}_{i=1}^{n}$ where $\tilde{h}_i = W_{\text{LT}_\mathcal{E}}$, for all $i$. Let $\tilde{H}$ denote the Hamiltonian obtained by applying $\text{LT}_{\tilde{\mathcal{E}}}$ to $H$ and note that the Level Transformation is such that $W_{\text{LT}_\mathcal{E}} = W_{\text{LT}_{\tilde{\mathcal{E}}}}$. To model this Level Transformation using thermal operations, let:

$$\tilde{H}_T = H \otimes |0\rangle\langle 0| + \tilde{H} \otimes |1\rangle\langle 1|,$$

and its work cost under thermal operations, $W_{H\to\tilde{H}}$, is given by the largest value of $W$ such that:

$$\left(\rho \otimes |0\rangle\langle 0| \otimes |0\rangle\langle 0|, \tilde{H}_T + H_W\right) \xrightarrow{\text{TO}} \left(\rho \otimes |1\rangle\langle 1| \otimes |1\rangle\langle 1|, \tilde{H}_T + H_W\right).$$

It can easily be seen that $W_{H\to\tilde{H}} = W_{\text{LT}_{\tilde{\mathcal{E}}}}$ and $W_{H\to\tilde{H}} \leq W_{H\to H'}$. Hence the result follows. □

Note that this result implies that implementing the effect of a Level Transformation using a switch qubit and thermal operations, can be more cost-effective (in terms of work required to make the transformation deterministically) then performing a Level Transformation itself.

## 6.3 Implementing allowed transformations using coarse operations

We now turn to showing how transitions between states that can be performed using thermal operations, can also be performed using the restricted set of coarse operations. Deriving this will be split into two parts. Firstly, we shall consider transformations between states with the same $\beta$-ordering and prove that they can be accomplished using only Partial Level Thermalizations that act on 2 energy levels at a time. Secondly, we show how to tackle states with different $\beta$-orderings, manipulating the initial and final states to form states with the same $\beta$-order. Two protocols will be presented for doing this, one making use of Exact Points Flows and the other Approximate Points Flows.

### 6.3.1 States with the same ordering

The action of Partial Level Thermalizations applied to 2 energy levels at a time, is analogous to that of the T-transforms introduced in Definition 25. Modifying Lemma 12 provides us with a protocol for converting between states with the same $\beta$-order under coarse operations:

**Theorem 20.** *Suppose that $\rho$ and $\sigma$ are diagonal states of an $n$-level system with Hamiltonian $H_S = \sum_{i=1}^{n} E_i|i\rangle\langle i|$ such that:*

1. *$\rho$ and $\sigma$ have the same $\beta$-order.*
2. *$\rho$ thermo-majorizes $\sigma$.*

*Then $\rho$ can be converted into $\sigma$ using at most $n-1$ Partial Level Thermalizations that each act on 2 energy levels.*

*Proof.* The aim is to construct a protocol consisting of such PLTs that converts $\rho$ into $\sigma$. To do this, we perform a sequence of PLTs. Each PLT adjusts the gradients of two line-segments of the thermo-majorization curve of $\rho$ until one of them matches the gradient of the corresponding segment on $\sigma$. By picking the segments of $\rho$ such that one has gradient strictly greater than the corresponding segment on $\sigma$ and one has gradient strictly less than the corresponding segment on $\sigma$, this can always be done. Once all of the gradients have been matched, $\rho$ has been converted into $\sigma$. The full details of the protocol are below and an illustration is given in Figure 6.6.

Let $\{\eta_i\}_{i=1}^n$ be the $\beta$-ordered eigenvalues of $\rho$, $\{\zeta_i\}_{i=1}^n$ be the $\beta$-ordered eigenvalues of $\sigma$ and $\{E_i\}_{i=1}^n$ be the $\beta$-ordered energy-eigenvalues of $H_S$. Hence we have:

$$\eta_1 e^{\beta E_1} \geq \cdots \geq \eta_n e^{\beta E_n},$$

and

$$\zeta_1 e^{\beta E_1} \geq \cdots \geq \zeta_n e^{\beta E_n}.$$

Given that $\rho$ and $\sigma$ have the same $\beta$-order, $\rho$ majorizes $\sigma$ if and only if:

$$\sum_{i=1}^{m} \eta_i \geq \sum_{i=1}^{m} \zeta_i, \quad \forall m. \tag{6.13}$$

Let $j$ be the largest index such that $\eta_j e^{\beta E_j} > \zeta_j e^{\beta E_j}$ and $k$ be the smallest index larger than $j$ such that $\eta_k e^{\beta E_k} < \zeta_k e^{\beta E_k}$. This picks the segments we shall apply the PLT to. Then:

$$\eta_j e^{\beta E_j} > \zeta_j e^{\beta E_j} \geq \zeta_k e^{\beta E_k} > \eta_k e^{\beta E_k}, \tag{6.14}$$

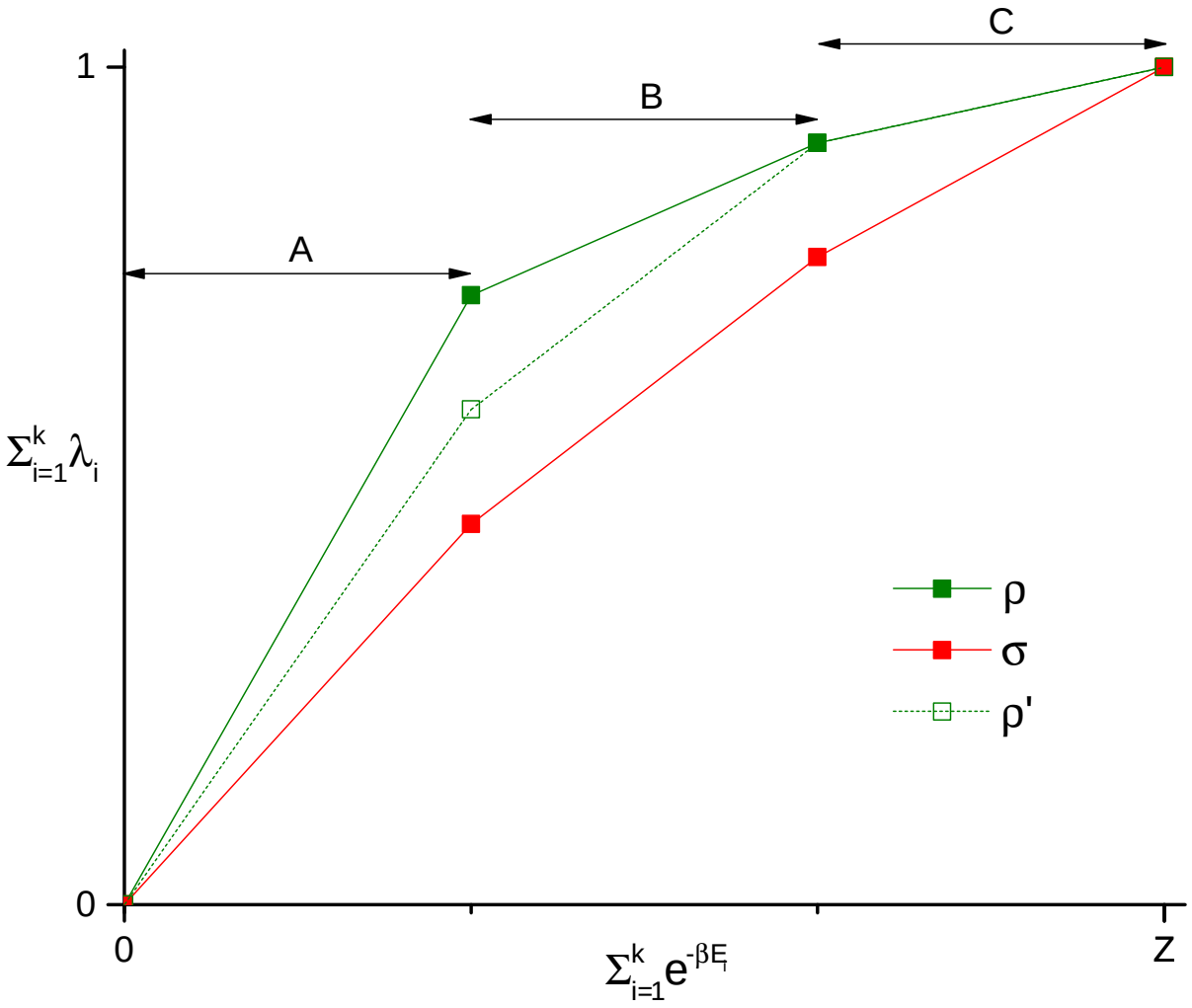



Figure 6.6: *Protocol for states with the same $\beta$-order.* The gradient of $\rho$ on segment $A$ is greater than that of $\sigma$ on the same segment. Similarly, the gradient of $\sigma$ is greater on segment $B$ than that of $\rho$. Hence, the first 2-level PLT is applied across segments $A$ and $B$. In this example, this leads to $\rho'$ such that the gradient of $\sigma$ and $\rho'$ are equal on segment $B$. A second run of the protocol matches the curves exactly.

and note that $\eta_i = \zeta_i$ for $j < i < k$.

We now determine the amount that we need to thermalize by in order to transform the gradient of one of the segments of $\rho$ to that of $\sigma$. Define $\lambda_1$ to be such that:

$$(1-\lambda_1)\,\eta_j e^{\beta E_j} + \frac{\lambda_1\,(\eta_j+\eta_k)}{e^{-\beta E_j}+e^{-\beta E_k}} = \zeta_j e^{\beta E_j},$$

and $\lambda_2$ to be such that:

$$(1-\lambda_2)\,\eta_k e^{\beta E_k} + \frac{\lambda_2\,(\eta_j+\eta_k)}{e^{-\beta E_j}+e^{-\beta E_k}} = \zeta_k e^{\beta E_k}.$$

Note that:

$$\eta_j e^{\beta E_j} \geq \frac{\eta_j+\eta_k}{e^{-\beta E_j}+e^{-\beta E_k}} \geq \eta_k e^{\beta E_k},$$

and hence $\lambda_1, \lambda_2 \geq 0$. Also, at least one of $\frac{(\eta_j+\eta_k)}{e^{-\beta E_j}+e^{-\beta E_k}} \leq \zeta_j e^{\beta E_j}$ or $\frac{(\eta_j+\eta_k)}{e^{-\beta E_j}+e^{-\beta E_k}} \geq \zeta_k e^{\beta E_k}$ holds as $\zeta_j e^{\beta E_j} \geq \zeta_k e^{\beta E_k}$. Hence, at least one of $\lambda_1$ and $\lambda_2$ must lie in the interval $[0,1]$. Let:

$$\lambda = \min\{\lambda_1, \lambda_2\}.$$

Let $\rho'$ be the state formed by applying the 2-level Partial Level Thermalization $\mathrm{PLT}_{\{j,k\}}(\lambda)$ to $\rho$. Note that $\rho'$ has the same $\beta$-order as $\rho$ and $\sigma$. To see this, let $\{\eta'_i\}_{i=1}^n$ be the eigenvalues of $\rho'$, listed according to the $\beta$ ordering of $\rho$. Then:

$$\begin{aligned}
\eta'_i &= \eta_i, \quad \text{for } 1 \leq i < j,\\
\eta'_i &= \zeta_i, \quad \text{for } j < i < k,\\
\eta'_i &= \eta_i, \quad \text{for } k < i \leq n,
\end{aligned}$$

as the Partial Level Thermalization does not change the occupation probabilities associated with $i \notin \{j,k\}$. Without loss of generality, suppose $\lambda = \lambda_1$ . Then $\eta'_j = \zeta_j$ and, using Eq. (6.14) where appropriate, it is easy to see that:

$$\eta'_i e^{\beta E_i} \geq \eta'_{i+1} e^{\beta E_{i+1}}, \quad \text{for both } i \in \{1,\ldots,k-2\} \text{ and } i \in \{k,\ldots,n\}.$$

To see that $\eta'_{k-1} e^{\beta E_{k-1}} \geq \eta'_k e^{\beta E_k}$, note that:

$$\eta'_k e^{\beta E_k} = (1-\lambda_1)\,\eta_k e^{\beta E_k} + \frac{\lambda_1\,(\eta_j+\eta_k)}{e^{-\beta E_j}+e^{-\beta E_k}} \leq \zeta_k e^{\beta E_k} \leq \zeta_{k-1} e^{\beta E_{k-1}} = \eta'_{k-1} e^{\beta E_{k-1}}.$$

Hence, the $\beta$-order of $\rho'$ is the same as $\rho$.

As Partial Level Thermalization is a thermal operation, $\rho$ thermo-majorizes $\rho'$. Similarly, $\rho'$ thermo-majorizes $\sigma$. To see this, it suffices to show that Eq. (6.13) still holds if we replace

$\rho$ with $\rho'$. As $\eta_j + \eta_k = \eta'_j + \eta'_k$, this obviously holds for $m < j$ and $m \geq k$. By observing that $\eta'_j \geq \zeta_j$ the remaining cases follow.

Applying the procedure once, sets at least one of the occupation probabilities to that of $\sigma$. Hence, by repeating the procedure at most $n-1$ times, starting each iteration with the output of the previous Partial Level Thermalization, we obtain $\sigma$. □

### 6.3.2 States with different ordering

Whilst Partial Level Thermalizations alone are not enough to perform a transformation allowed under thermal operations, by combining them with Level Transformations and the ability to append a single ancillary system with known Hamiltonian in the Gibbs state, they become more powerful. Indeed, they can be used to perform any transition between diagonal states allowed under thermal operations without the need to expend any work. This is captured and proven in the following theorem:

**Theorem 21.** *Suppose that $\rho$ and $\sigma$ are diagonal states of an $n$-level system with Hamiltonian $H_S = \sum_{i=1}^{n} E_i |i\rangle\langle i|$ such that:*

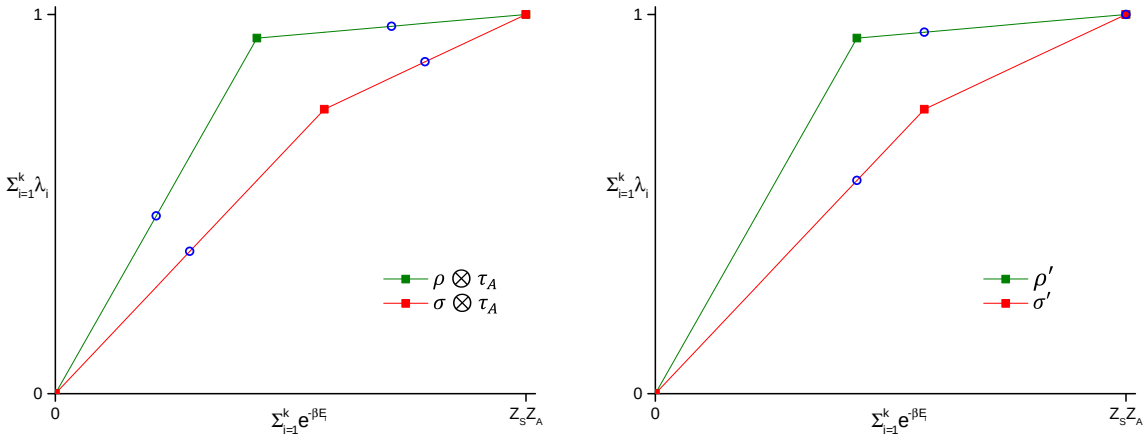
1. *$\rho$ thermo-majorizes $\sigma$.*

*Then $\rho$ can be converted into $\sigma$ using coarse operations in a protocol that has work cost arbitrarily close to zero with probability arbitrarily close to one.*

*Proof.* To prove this, we give a protocol consisting only of: adding (and eventually discarding) an ancilla thermal qubit, Exact Points Flow protocols (as introduced in Definition 30) and Partial Level Thermalizations. By Lemma 17, Exact Points Flows have work cost arbitrarily highly peaked around zero while both Partial Level Thermalizations and using the ancilla qubit cost no work.

Let $(\tau_A, H_A)$ denote the known ancilla qubit allowed under Operation 1 of coarse operations. The protocol then runs as follows:

$$\begin{aligned}
(\rho, H_S) &\longrightarrow (\rho \otimes \tau_A, H_S + H_A)\,, \\
&\xrightarrow{\text{EPF}} \left(\rho', H'_{SA}\right), \\
&\xrightarrow{\text{PLT}} \left(\sigma', H'_{SA}\right), \\
&\xrightarrow{\text{EPF}} (\sigma \otimes \tau_A, H_S + H_A)\,, \\
&\longrightarrow (\sigma, H_S)\,.
\end{aligned}$$




(a) Effect of appending the ancilla.

(b) The states $\rho'$ and $\sigma'$.

Figure 6.7: *$\beta$-order change with Exact Points Flows.* Figure (a) shows $\rho \otimes \tau_A$ and $\sigma \otimes \tau_A$ with the non-elbow points introduced represented by blue circles. Using EPFs, the non-elbow points of $\rho \otimes \tau_A$ and $\sigma \otimes \tau_A$ can be moved so that they lie horizontally aligned with the elbows of $\sigma$ and $\rho$ respectively. These states, $\rho'$ and $\sigma'$, are shown in Figure (b). As $\rho'$ and $\sigma'$ have the same $\beta$-ordering and $\rho'$ thermo-majorizes $\sigma'$, coarse operations can be used to convert $\rho'$ into $\sigma'$.

Here $(\rho', H'_{SA})$ is a system with the same thermo-majorization curve as $(\rho \otimes \tau_A, H_S + H_A)$. However, the non-elbow points have been moved (potentially while sending energy levels to infinity) so that on the thermo-majorization diagram, they are vertically inline with the elbows (including the point $(Z, 1)$) of $(\sigma \otimes \tau_A, H_S + H_A)$. This transformation can be performed using EPFs and PITRs.

Similarly, $(\sigma', H'_{SA})$ is defined to be a system with the same thermo-majorization curve as $(\sigma \otimes \tau_A, H_S + H_A)$ but with the non-elbow points moved to lie vertically inline with the elbows of $(\rho \otimes \tau_A, H_S + H_A)$ and at $(Z, 1)$. Again, this transformation can be performed (and reversed) using the Points Flow protocol.

Note that by construction, $(\rho', H'_{SA})$ has the same $\beta$-ordering as $(\sigma', H'_{SA})$ and as $\rho$ thermo-majorizes $\sigma$, $(\rho', H'_{SA})$ thermo-majorizes $(\sigma', H'_{SA})$. Hence, by Theorem 20, it is possible to transform $(\rho', H'_{SA})$ into $(\sigma', H'_{SA})$ using Partial Level Thermalizations. □

The states of the protocol are illustrated in Figure 6.7.

The protocol described in the above theorem potentially requires that an energy level be raised to infinite energy. While this can be done at no work cost (provided it is performed infinitely slowly during the Exact Points Flow protocol), note that such a transition is not

required if the thermo-majorization curves of $\rho$ and $\sigma$ do not cross.

**Theorem 22.** *Suppose that $\rho$ and $\sigma$ are diagonal states of an $n$-level system with Hamiltonian $H_S = \sum_{i=1}^{n} E_i |i\rangle\langle i|$ such that:*

1. *$\rho$ thermo-majorizes $\sigma$.*

2. *The thermo-majorization curves of $\rho$ and $\sigma$ meet only at $(0,0)$ and $(Z_S, 1)$.*

*Then $\rho$ can be converted into $\sigma$ using coarse operations in a protocol that has work cost arbitrarily close to zero with probability arbitrarily close to one and without the need to raise an energy level to infinity.*

*Proof.* Here we sketch how to modify the protocol given in Theorem 21 to avoid needing to raise an energy level to infinity. The new protocol runs as follows:

$$\begin{aligned}(\rho, H_S) &\longrightarrow (\rho \otimes \tau_A, H_S + H_A)\,, \\ &\overset{\substack{\text{APF}\\ \text{PLT}}}{\longrightarrow} (\tilde{\rho}, H_S + H_A)\,, \\ &\overset{\text{PLT}}{\longrightarrow} (\sigma \otimes \tau_A, H_S + H_A)\,, \\ &\longrightarrow (\sigma, H_S)\,.\end{aligned}$$

Here $(\tilde{\rho}, H_S + H_A)$ is a system with a thermo-majorization curve such that each one of its points (both elbows and non-elbows) are vertically aligned with the points of $(\sigma \otimes \tau_A, H_S + H_A)$.

To create $(\tilde{\rho}, H_S + H_A)$, we use the following process, illustrated in Figure 6.8:

1. Using Approximate Points Flows, adjust the points of $\rho \otimes \tau_A$ to form $\rho'$ which has non-elbow points vertically aligned with the elbows of $\sigma \otimes \tau_A$. There are $n-1$ such points. As the thermo-majorization curves of $\rho \otimes \tau_A$ and $\sigma \otimes \tau_A$ touch only at $(0,0)$ and at $(Z_S Z_A, 1)$, the APF can be chosen such that $\rho'$ has the desired alignment, thermo-majorizes $\sigma \otimes \tau_A$ and such that the thermo-majorization curves of $\rho'$ inherits these properties.

2. For each vertically aligned point $i \in \{1, \ldots n-1\}$ on $\rho'$, consider the number of points (both elbows and non-elbows) to the left of it on its thermo-majorization curve. Call this number $r_i$. Compare this quantity to the number of points to the left of the associated vertically aligned point on the thermo-majorization curve of $\sigma \otimes \tau_A$. Call this number $s_i$. If:

(a) $r_i < s_i$: Move the point slightly to the right of its aligned location using a PITR.

(b) $r_i > s_i$: Move the point slightly to the left of its aligned location using a PITR.

(c) $r_i = s_i$: Leave the point where it is.

These PITRs result in a state $\rho''$ with the same thermo-majorization curve as $\rho'$.

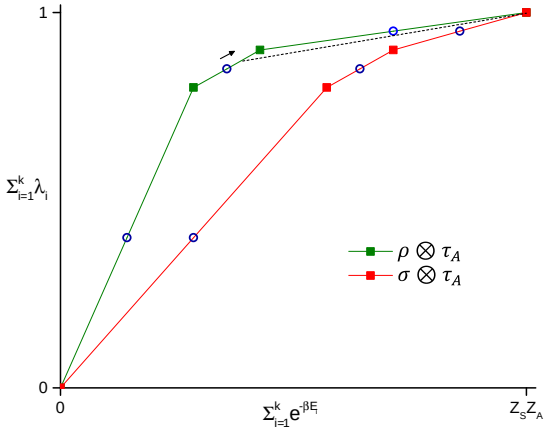
3. Defining $i = 0$ to be the point $(0, 0)$ and $i = n$ to be the point $(Z_S Z_A, 1)$, for each $i \in \{1, \dots n\}$ thermalize $\rho''$ over the the interval between points $i - 1$ and $i$ using PLTs. This results in a state $\rho'''$ which has elbows almost vertically aligned with the elbows of $\sigma \otimes \tau_A$. Provided the movements due to PITRs in Step 2 were chosen to be sufficiently small, as $\rho''$ thermo-majorizes $\sigma \otimes \tau_A$ and their thermo-majorization curves touch only at $(0, 0)$ and $(Z_S Z_A, 1)$, $\rho'''$ inherits the same properties.

4. Using Approximate Points Flows, adjust the points of $\rho'''$ to form $\tilde{\rho}$ as defined above. The last time an APF is applied around an elbow, it should be done in such a way that after the operation, an elbow is precisely vertically aligned with that of $\sigma \otimes \tau_A$. The displacements applied in Step 2 enable this to take place. As $\rho'''$ thermo-majorizes $\sigma \otimes \tau_A$ and their thermo-majorization curves touch only at (0,0) and $(Z_S Z_A, 1)$, $\tilde{\rho}$ again inherits these properties.

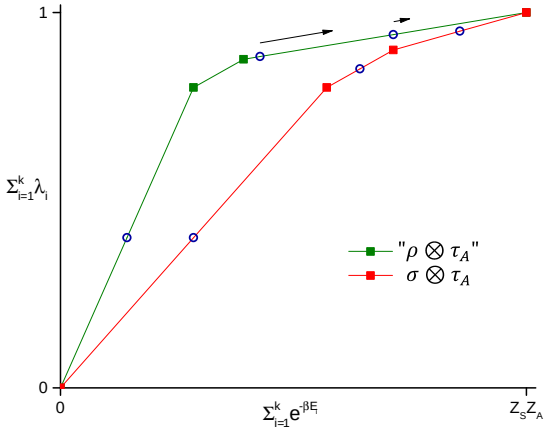
Due to the fact that the protocol uses Approximate Points Flows rather than Exact Points Flows, there is no need to raise an energy level to infinity.

As $(\tilde{\rho}, H_S + H_A)$ thermo-majorizes $\sigma \otimes \tau_A$ and they have the same $\beta$-ordering, the transformation can now be completed using Partial Level Thermalizations as described in Theorem 20. □

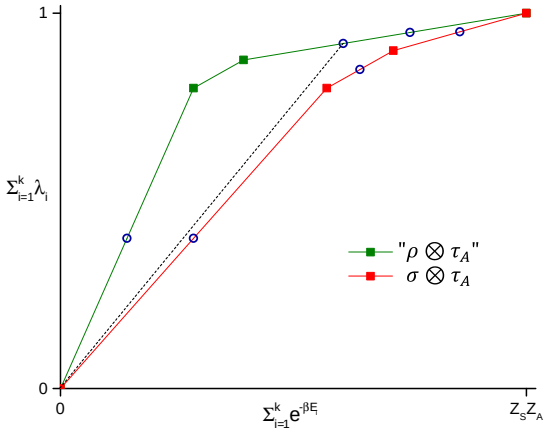
Note that there are scenarios in which the restriction that the curves touch only at $(0, 0)$ and $(Z_S, 1)$ in the above theorem can be relaxed slightly to demanding that the curves touch only at $(0, 0)$ and on the line $y = 1$. For example, this is the case if $|W_{\rho \to \tau_S}| > |W_{\sigma \to \tau_S}|$, where $\tau_S$ is the Gibbs state of the system's Hamiltonian, i.e. it is possible to extract strictly more work deterministically from $\rho$ than from $\sigma$. In general, allowing the curves to touch at $y = 1$ makes the theorem more relevant for situations where we wish to model a change of Hamiltonian or include a work storage system as per Eq. (4.34).



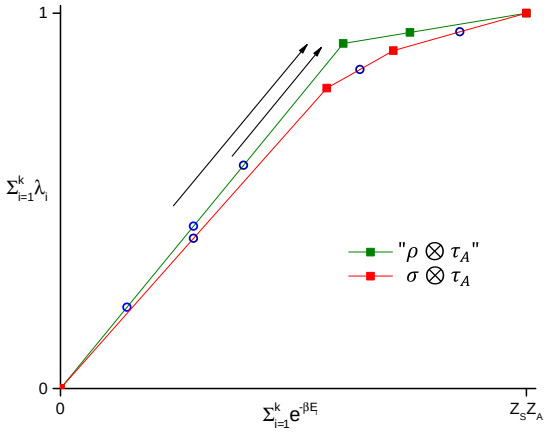

(a) Initial state. Step 1: APFs.



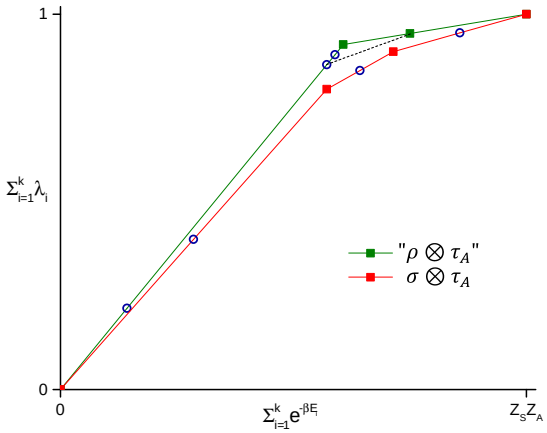

(b) Step 2: PITRs.



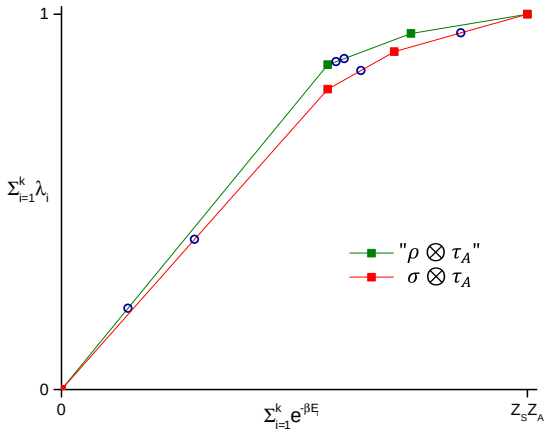

(c) Step 3: PLTs.



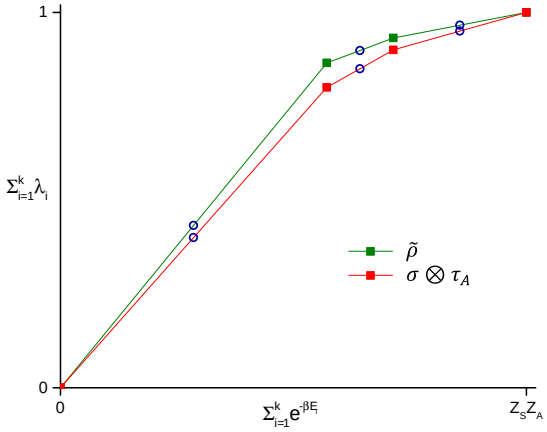

(d) Step 4: APFs.




(e) Step 4: APFs continued.




(f) Step 4: APFs continued.




(g) Final state with the same $\beta$-ordering as $\sigma \otimes \tau_A$.

Figure 6.8: *$\beta$-order change with Approximate Points Flows.* In Step 1 of the protocol, APFs are performed so that the non-elbows of $\rho \otimes \tau_A$ can be horizontally aligned with the elbows of $\sigma \otimes \tau_A$. This transforms Figure (a) into Figure (b). In Step 2, PITRs are used to slightly misaligned the points in anticipation of Step 4. In this example they are misaligned to the right resulting in Figure (c). Next, PLTs are used in Step 3 to generate elbows (almost) horizontally aligned with those of $\sigma \otimes \tau_A$. This leads to Figure (d). Finally, in Step 4 APFs are applied to exactly match the elbows together with the non-elbow points. This is detailed in Figures (e) and (f). The result of the protocol is shown in Figure (g).

## 6.4 Summary

In this chapter, we have shown that the greatly restricted set of operations defined by coarse operations are sufficient for performing those transitions between diagonal states that are possible using thermal operations. Furthermore, we have given methods for constructing explicit protocols to implement these transformations. As coarse operations require comparatively little control over the system and heat bath, these results and protocols serve to bring thermodynamical transformations that have never before been performed in a laboratory into a regime accessible by experiment.

In [2], it was shown that the optimal amount of work distillable from a state (both on average and in the worst-case) can be quantified in terms of the Level Transformations used through Eq. (6.5). Under coarse operations, Level Transformations can also be used to capture the work cost/yield of a generic transformation, $W_{\rho\to\sigma}$. Given two states, $\rho$ and $\sigma$, if one first performs the Level Transformation $\mathrm{LT}_{\mathcal{E}}$, with $\mathcal{E} = \{h_i = W_{\rho\to\sigma}\}_{i=1}^{n}$, to $(\rho, H_S)$, by definition one is left with a system that thermo-majorizes $(\sigma, H_S)$. This system can now be converted into $(\sigma, H_S)$ using the protocol given in Theorem 21 and the work cost of the Level Transformation is $W_{\rho\to\sigma}$. The advantage of modeling work in this way, rather than making use of the work storage system given by Eq. (4.28), lies in the fact that one does not need to precisely engineer a Hamiltonian to mimic such a system. This would add an extra layer of complexity to an experimental implementation of thermal operations.

By allowing for the presence and manipulation of a catalyst, coarse operations can also be used to implement the transformations possible under the catalytic thermal operations mentioned briefly in Section 4.3.1. As shown in [24], provided the initial state is diagonal in the energy eigenbasis, a catalyst that is diagonal with respect to its Hamiltonian is sufficient for performing those transitions allowed by the generalized free energies. Hence, the protocols defined here can equally well be applied to the joint state of system and catalyst. There is however, a caveat. The dimension of the catalyst used to assist with the transformation can be much larger than the dimension of the system and hence manipulating it may become difficult experimentally. Work towards characterizing the transformations allowed using catalysts of bounded dimension can be found in [132].

Coarse operations however, do not allow one to implement all transformation between states which are possible under thermal operations. The problem once again revolves around states

with coherence. Here, in addition to the issues regarding non-block-diagonal target states, challenges arise surrounding initial states containing coherence. Indeed, decohering in the energy eigenbasis, as per Eq. (4.23), is not possible under coarse operations in general. As protecting against decoherence is usually of grave concern when dealing with quantum systems, one could readily add the operation of decohering to the set of allowed operations without drastically affecting their applicability to experiment. Overall however, as necessary and sufficient conditions for control of quantum coherences are not known for thermal operations, it is difficult to envisage what additional actions should be added to coarse operations to enable them to implement the full range of thermodynamical processes.

There remains much work to be done in understanding, developing and exploiting the laws of thermodynamics at the nano-scale and beyond. The resource theory approach captured by thermal operations provides a path towards this goal. As yet however, its results have not been subjected to experimental test. The results in this chapter go someway to rectifying this.

# Part III

# Appendices

# Appendix A

# Proofs for state exclusion

This appendix contains proofs for results regarding state exclusion that were stated or mentioned in the main text.

## A.1 A proof of the necessary condition for conclusive state discrimination

Here we give a derivation of Corollary 1 from Theorem 4.

**Corollary.** *Conclusive state discrimination on the set $\mathcal{P} = \{\rho_i\}_{i=1}^k$ is possible only if $\mathcal{P}$ is an orthogonal set.*

*Proof.* For $\mathcal{P} = \{\rho_i\}_{i=1}^k$, define:

$$\hat{\rho}_j = \frac{1}{k-1} \sum_{i \neq j} \rho_i.$$

Let $j \neq l$ and consider:

$$A = \frac{1}{k-1} \sum_{r \neq j,l} \rho_r.$$

We first show that $F(\hat{\rho}_j, \hat{\rho}_l) \geq F(\hat{\rho}_j, A)$. Consider:

$$\begin{aligned} F(\hat{\rho}_j, A) &= \mathrm{Tr}\left[\sqrt{\sqrt{\hat{\rho}_j} A \sqrt{\hat{\rho}_j}}\right], \\ &\leq \mathrm{Tr}\left[\sqrt{\sqrt{\hat{\rho}_j} \hat{\rho}_l \sqrt{\hat{\rho}_j}}\right], \\ &= F(\hat{\rho}_j, \hat{\rho}_l). \end{aligned}$$

The inequality follows from the following facts:

1. It can be easily seen from the definitions that $A \leq \hat{\rho}_l$.

2. If $B \geq C$ then $D^\dagger B D \geq D^\dagger C D$, $\forall D$. Hence:

$$\sqrt{\hat{\rho}_j} A \sqrt{\hat{\rho}_j} \leq \sqrt{\hat{\rho}_j} \hat{\rho}_l \sqrt{\hat{\rho}_j}.$$

3. The square root function is operator monotone, so:

$$\sqrt{\sqrt{\hat{\rho}_j} A \sqrt{\hat{\rho}_j}} \leq \sqrt{\sqrt{\hat{\rho}_j} \hat{\rho}_l \sqrt{\hat{\rho}_j}}.$$

4. The trace function is operator monotone and so finally:

$$\mathrm{Tr}\left[\sqrt{\sqrt{\hat{\rho}_j} A \sqrt{\hat{\rho}_j}}\right] \leq \mathrm{Tr}\left[\sqrt{\sqrt{\hat{\rho}_j} \hat{\rho}_l \sqrt{\hat{\rho}_j}}\right].$$

Using a similar argument to the above, it is possible to show that:

$$F\left(\hat{\rho}_j, A\right) \geq F\left(A, A\right) = \frac{k-2}{k-1}.$$

If $\rho_j$, $\rho_l$ and $A$ are pairwise orthogonal, then $\hat{\rho}_j$ and $\hat{\rho}_l$ commute and are simultaneously diagonalizable. This means that:

$$\begin{aligned} F\left(\hat{\rho}_j, \hat{\rho}_l\right) &= \left|\left|\sqrt{\hat{\rho}_j}\sqrt{\hat{\rho}_l}\right|\right|_{\mathrm{tr}}, \\ &= \left|\left|A\right|\right|_{\mathrm{tr}}, \\ &= F\left(A, A\right), \\ &= \frac{k-2}{k-1}. \end{aligned}$$

Now suppose that $\rho_j$ and $A$ are not orthogonal. We take $\{a_r\}$ to be the eigenvalues and $\{|v_r\rangle\}$ to be the eigenvectors of $\sqrt{A}$, so:

$$\begin{aligned} F\left(\hat{\rho}_l, A\right) &\geq \mathrm{Tr}\left[\sqrt{\hat{\rho}_l}\sqrt{A}\right], \\ &= \sum_r a_r \langle v_r | \sqrt{\hat{\rho}_l} | v_r \rangle. \end{aligned}$$

We know that $\sqrt{\hat{\rho}_l} \geq \sqrt{A}$ and hence:

$$\langle v_r | \sqrt{\hat{\rho}_l} | v_r \rangle \geq a_r, \quad \forall r.$$

As $\rho_j$ and $A$ are not orthogonal:

$$\sum_r \langle v_r | \sqrt{\hat{\rho}_l} | v_r \rangle > \sum_r a_r,$$

and there must exist some $r$ such that:

$$\langle v_r|\sqrt{\hat{\rho}_l}|v_r\rangle > a_r.$$

Hence:

$$\begin{aligned} F\left(\hat{\rho}_l, A\right) &\geq \sum_r a_r \langle v_r|\sqrt{\hat{\rho}_l}|v_r\rangle, \\ &> \sum_r a_r^2, \\ &= \mathrm{Tr}\left[A\right], \\ &= \frac{k-2}{k-1}. \end{aligned}$$

So $F\left(\hat{\rho}_j, \hat{\rho}_l\right) = \frac{k-2}{k-1}$, $\forall l \neq j$, if and only if $\mathcal{P}$ is an orthogonal set.

By Theorem 4, for conclusive $(m-1)$-state exclusion (and hence conclusive state discrimination) to be possible, we require that:

$$\sum_{j\neq l=1}^{k} F\left(\hat{\rho}_j, \hat{\rho}_l\right) = k\left(k-2\right),$$

which implies that $\mathcal{P}$ must be an orthogonal set. □

## A.2 Optimality of projective measurements

Here, we discuss the conditions for which a projective measurement is optimal for performing single state exclusion on a set of pure states.

Let $\mathcal{M}$ and $N$ be optimal solutions to the primal and dual state exclusion SDPs given in Eqs. (2.21) and (2.22). By Proposition 2 (complementary slackness), they are such that:

$$\begin{aligned} &\begin{pmatrix} NM_1 & & \\ & \ddots & \\ & & NM_k \end{pmatrix} = \begin{pmatrix} \rho_1 M_1 & & \\ & \ddots & \\ & & \rho_k M_k \end{pmatrix}, \\ &\Rightarrow \left(N - \rho_i\right) M_i = 0, \quad \forall i, \\ &\Rightarrow M_i \text{ lies in the nullspace of } \left(N - \rho_i\right), \\ &\Rightarrow \mathrm{rank}(M_i) \leq \dim\left[\text{nullspace}\left(N - \rho_i\right)\right]. \end{aligned}$$

Now, for two operators $A$ and $B$:

$$\dim\left[\text{nullspace}\left(A+B\right)\right] \leq \dim\left[\text{nullspace}\left(A\right)\right] + \mathrm{rank}\left(B\right).$$

So, if the $\rho_i = |\psi_i\rangle\langle\psi_i|$ are pure states and $N$ has full rank, the optimal measurement consists of orthogonal rank 1 projectors.

To see when $N$ has full rank, suppose there exists an $|x\rangle$ such that $N|x\rangle = 0$. Then:

$$\begin{aligned} & M_i \left(N - |\psi_i\rangle\langle\psi_i|\right)|x\rangle = 0, \quad \forall i, \\ \Rightarrow \qquad & M_i|\psi_i\rangle\langle\psi_i|x\rangle = 0, \quad \forall i. \end{aligned}$$

Suppose that $M_i|\psi_i\rangle \neq 0$, for all $i$. Then:

$$\langle\psi_i|x\rangle = 0, \quad \forall i.$$

So if:

- Conclusive exclusion is not possible for any measurement outcome: $M_i|\psi_i\rangle \neq 0$, for all $i$.
- The set $\mathcal{P} = \{|\psi_i\rangle\}_{i=1}^k$ is a linearly independent set.

then $N|x\rangle = 0$ implies that $|x\rangle = 0$ and hence $N$ has full rank and the optimal measurement for performing exclusion consists of orthogonal rank 1 projectors.

# Appendix B

# SDP formulations

This Appendix contains the details behind the derivation of some of the dual SDPs given in Chapter 2.

## B.1 The unambiguous state exclusion SDP

Here we give the derivation of Eq. (2.37).

Comparing Eq. (2.36) with Eq. (1.2), we see that here:

- $A$ is a $kd$ by $kd$ block- diagonal matrix with each $d$ by $d$ block containing $\sum_{j=1}^{k} \tilde{\rho}_j$:

$$A = \begin{pmatrix} \sum_{j=1}^{k} \tilde{\rho}_j & & \\ & \ddots & \\ & & \sum_{j=1}^{k} \tilde{\rho}_j \end{pmatrix}. \tag{B.1}$$

- $B$ is a $(d+k)$ by $(d+k)$ matrix with the top left $d$ by $d$ block being an identity matrix and all other elements being 0:

$$B = \begin{pmatrix} \mathbb{I} & 0 \\ 0 & 0 \end{pmatrix}. \tag{B.2}$$

- $X$, the variable matrix, is a $kd$ by $kd$ block-diagonal matrix where we label each $d$ by $d$ block-diagonal by $M_i$:

$$X = \begin{pmatrix} M_1 & & \\ & \ddots & \\ & & M_k \end{pmatrix}. \tag{B.3}$$

- $Y$ is a $(d+k)$ by $(d+k)$ matrix whose top left $d$ by $d$ block we call $N$ and the remaining $k$ diagonal elements we label by $a_i$.

$$Y = \begin{pmatrix} N & & & \\ & a_1 & & \\ & & \ddots & \\ & & & a_k \end{pmatrix}. \tag{B.4}$$

- The map $\Phi$ is given by:

$$\Phi(X) = \begin{pmatrix} \sum_{i=1}^{k} M_i & & & \\ & \mathrm{Tr}\left[\tilde{\rho}_1 M_1\right] & & \\ & & \ddots & \\ & & & \mathrm{Tr}\left[\tilde{\rho}_k M_k\right] \end{pmatrix}. \tag{B.5}$$

Using Eq. (1.4), we see that $\Phi^*$ must satisfy:

$$\mathrm{Tr}\left[N\sum_{i=1}^{k} M_i\right] + \sum_{i=1}^{k} a_i \mathrm{Tr}\left[\tilde{\rho}_i M_i\right] = \mathrm{Tr}\left[\begin{pmatrix} M_1 & & \\ & \ddots & \\ & & M_k \end{pmatrix} \Phi^* \left[\begin{pmatrix} N & & & \\ & a_1 & & \\ & & \ddots & \\ & & & a_k \end{pmatrix}\right]\right], \tag{B.6}$$

and hence $\Phi^*(Y)$ produces a $kd$ by $kd$ block-diagonal matrix:

$$\Phi^*(Y) = \begin{pmatrix} N + a_1\tilde{\rho}_1 & & \\ & \ddots & \\ & & N + a_k\tilde{\rho}_k \end{pmatrix}. \tag{B.7}$$

Substituting these elements into Eq. (1.3) and noting that we are maximizing in the primal problem, we find that the dual problem is given by Eq. (2.37).

## B.2 The worst-case error state exclusion SDP

Here we give the derivation of Eq. (2.40).

Comparing Eq. (2.39) with Eq. (1.2), we see that here:

- $A$ is a $(kd+1)$ by $(kd+1)$ matrix with $A_{11}=1$ being the only non-zero element:

$$A = \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}. \tag{B.8}$$

- $B$ is a $(d+k)$ by $(d+k)$ matrix where the bottom right $d$ by $d$ block is the identity matrix. All other elements are zero:

$$B = \begin{pmatrix} 0 & 0 \\ 0 & \mathbb{I} \end{pmatrix}. \tag{B.9}$$

- $X$, the variable matrix, is a $kd+1$ by $kd+1$ block-diagonal matrix where $X_{11}=\lambda$ and we label each subsequent $d$ by $d$ block-diagonal by $M_i$:

$$X = \begin{pmatrix} \lambda & & & \\ & M_1 & & \\ & & \ddots & \\ & & & M_k \end{pmatrix}. \tag{B.10}$$

- $Y$ is a $(d+k)$ by $(d+k)$ matrix whose bottom right $d$ by $d$ block we shall call $N$ and the remaining $k$ diagonal elements we label by $a_i$.

$$Y = \begin{pmatrix} a_1 & & & \\ & \ddots & & \\ & & a_k & \\ & & & N \end{pmatrix}. \tag{B.11}$$

- The map $\Phi$ is given by:

$$\Phi(X) = \begin{pmatrix} \lambda - \mathrm{Tr}\left[\tilde{\rho}_1 M_1\right] & & & \\ & \ddots & & \\ & & \lambda - \mathrm{Tr}\left[\tilde{\rho}_k M_k\right] & \\ & & & \sum_{i=1}^{k} M_i \end{pmatrix}. \tag{B.12}$$

Using Eq. (1.4), we see that $\Phi^*$ must satisfy:

$$\lambda \sum_{i=1}^{k} a_i - \sum_{i=1}^{k} a_i \mathrm{Tr}\left[\tilde{\rho}_i M_i\right] = \mathrm{Tr}\left[\begin{pmatrix} \lambda & & & \\ & M_1 & & \\ & & \ddots & \\ & & & M_k \end{pmatrix} \Phi^* \left[\begin{pmatrix} a_1 & & & \\ & \ddots & & \\ & & a_k & \\ & & & N \end{pmatrix}\right]\right], \quad \text{(B.13)}$$

and hence $\Phi^*(Y)$ produces a $kd$ by $kd$ block-diagonal matrix:

$$\Phi^*(Y) = \begin{pmatrix} \sum_{i=1}^{k} a_i & & & \\ & N - a_1\tilde{\rho}_1 & & \\ & & \ddots & \\ & & & N - a_k\tilde{\rho}_k \end{pmatrix}. \quad \text{(B.14)}$$

Substituting these elements into Eq. (1.3), we obtain Eq. (2.40).

## B.3 The measure of equal support compatibility SDP

Here we give the derivation of Eq. (2.53).

First we rewrite Eq.(2.52) so that it has the same structure as Eq. (1.2). This leads to:

$$\begin{aligned} \underset{\lambda,\{\lambda_i\}_{i=1}^{d}}{\text{Maximize:}} \ & \mathrm{Tr}\left[\begin{pmatrix} \lambda & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix} \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}\right]. \\ \text{Subject to:} \ & \lambda - \lambda_i \leq 0, \quad \forall i, \\ & \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix} \sum_{j=1}^{k} \rho_j \leq \rho_i, \quad \forall i, \\ & \lambda \geq 0, \\ & \lambda_i \geq 0, \quad \forall i. \end{aligned} \quad \text{(B.15)}$$

Comparing Eq. (B.15) with Eq. (1.2), we see that:

- $A$ is a $d+1$ by $d+1$ matrix:

$$A = \begin{pmatrix} 1 & & & \\ & 0 & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}. \tag{B.16}$$

- $B$ is a $d(k+1)$ by $d(k+1)$ matrix where the first $d$ entries on the diagonal are 0, and the remaining matrix is block-diagonal with the blocks given by $\rho_i$:

$$B = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & \rho_1 & & \\ & & & & \ddots & \\ & & & & & \rho_k \end{pmatrix}. \tag{B.17}$$

- $X$, the variable matrix, is a $d+1$ by $d+1$ matrix:

$$X = \begin{pmatrix} \lambda & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix}. \tag{B.18}$$

- $Y$ is a $d(k+1)$ by $d(k+1)$ matrix whose first $d$ entries on the diagonal we label by $\alpha_i$, and the remaining block-diagonal with the elements we denote by $M_i$:

$$Y = \begin{pmatrix} \alpha_1 & & & & & \\ & \ddots & & & & \\ & & \alpha_d & & & \\ & & & M_1 & & \\ & & & & \ddots & \\ & & & & & M_k \end{pmatrix}. \tag{B.19}$$

- The map $\Phi$ is given by:

$$\Phi(X) = \begin{pmatrix} \lambda - \lambda_1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda - \lambda_d & & & & & & \\ & & & \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix} \sum_{i=1}^{k} \rho_i & & & & & \\ & & & & \ddots & & & & \\ & & & & & \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix} \sum_{i=1}^{k} \rho_i & & \\ & & & & & & & & \end{pmatrix}. \tag{B.20}$$

Using Eq. (1.4), we see that $\Phi^*$ must satisfy:

$$\begin{aligned} &\sum_{i=1}^{d} \alpha_i (\lambda - \lambda_i) + \sum_{i=1}^{k} \mathrm{Tr}\left[ M_i \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{pmatrix} \sum_{j=1}^{k} \rho_j \right] \\ &\qquad = \mathrm{Tr}\left[ \begin{pmatrix} \lambda & & & \\ & \lambda_1 & & \\ & & \ddots & \\ & & & \lambda_d \end{pmatrix} \Phi^* \left( \begin{pmatrix} \alpha_1 & & & & & \\ & \ddots & & & & \\ & & \alpha_d & & & \\ & & & M_1 & & \\ & & & & \ddots & \\ & & & & & M_k \end{pmatrix} \right) \right], \end{aligned} \tag{B.21}$$

and hence $\Phi^*(Y)$ produces a $d+1$ by $d+1$ matrix:

$$\Phi^*(Y) = \begin{pmatrix} \sum_{i=1}^{d} \alpha_i & & \\ & \begin{pmatrix} -\alpha_1 & & \\ & \ddots & \\ & & -\alpha_d \end{pmatrix} + \sum_{i=1}^{k} \rho_i \sum_{j=1}^{k} M_i \end{pmatrix}. \tag{B.22}$$

If we now substitute these elements into Eq. (1.3), we obtain Eq. (2.53).

# Appendix C

# The tradeoff between the probability and the purity of a transition

In this appendix, we consider how $p^*$ varies if one supplies additional purity when attempting to convert $\rho$ into $\sigma$ using noisy operations, as discussed at the end of Chapter 5. Alternatively, one could attempt to extract extra purity during the process. Whilst characterizing the behavior of $p^*$ in general for thermal operations is an open question (though see [149] for recent progress), here we give the solution for qubit systems under noisy operations.

Consider two qubits: $\rho$ with ordered eigenvalues $\vec{\eta} = \{\eta_1, \eta_2\}$ and $\sigma$ with ordered eigenvalues $\vec{\zeta} = \{\zeta_1, \zeta_2\}$. For the transition:

$$\begin{aligned} \rho \otimes s_{|S|} &\xrightarrow{NO} \rho' = p\sigma + (1-p)\, X, && \text{if } S \leq 0, \\ \rho &\xrightarrow{NO} \rho' = p\sigma \otimes s_{|S|} + (1-p)\, X, && \text{if } S > 0, \end{aligned} \tag{C.1}$$

how does $p^*$ behave as a function of $S$? Note that using Theorem 18, $p^*(0)$ is given by $\min\left\{\frac{\eta_1}{\zeta_1}, 1\right\}$. For $S \leq S_{\rho\to\sigma}$, by definition we have that $p^*(S) = 1$. So as to investigate the behavior of the function at $S = 0$, in what follows we shall assume $\eta_1 < \zeta_1$ and hence $S_{\rho\to\sigma} < 0$.

First take $S \leq 0$ and for simplicity, assume it can be written as $S = -\log_2 \frac{d}{j}$. Then:

$$\begin{aligned} \rho \otimes s_{|S|} &= \operatorname{diag}\Big(\underbrace{\tfrac{\eta_1}{j}, \ldots, \tfrac{\eta_1}{j}}_{j}, \underbrace{\tfrac{\eta_2}{j}, \ldots, \tfrac{\eta_2}{j}}_{j}, \underbrace{0, \ldots, 0}_{2(d-j)}\Big), \\ \sigma \otimes \frac{\mathbb{I}}{d} &= \operatorname{diag}\Big(\underbrace{\tfrac{\zeta_1}{d}, \ldots, \tfrac{\zeta_1}{d}}_{d}, \underbrace{\tfrac{\zeta_2}{d}, \ldots, \tfrac{\zeta_2}{d}}_{d}\Big). \end{aligned} \tag{C.2}$$

We now use Theorem 18 together with the fact that $p^*(S)$ will occur at an 'elbow' of $\sigma$ (which is equivalent to $\sigma \otimes \frac{\mathbb{I}}{d}$ under noisy operations). As $S_{\rho\to\sigma} < S$, we need only consider the elbow

at $l = d$. Thus:

$$p^*(S) = \frac{V_d\left(\rho \otimes s_{|S|}\right)}{V_d\left(\sigma \otimes \frac{\mathbb{I}}{d}\right)} = \frac{\eta_1 + \frac{d-j}{j}\eta_2}{\zeta_1}, \quad S_{\rho\to\sigma} < -\log_2 \frac{d}{j} \leq 0. \tag{C.3}$$

This can be rearranged to give:

$$p^*(S) = \left(2 - 2^{-S}\right) p^*(0) + \frac{2^{-S} - 1}{\zeta_1}, \qquad S_{\rho\to\sigma} < S \leq 0. \tag{C.4}$$

Now take $S \geq 0$ and assume it can be written as $S = \log_2 \frac{d}{j}$. Then:

$$\begin{aligned} \rho \otimes \frac{\mathbb{I}}{d} &= \mathrm{diag}\Big(\underbrace{\tfrac{\eta_1}{d}, \ldots, \tfrac{\eta_1}{d}}_{d}, \underbrace{\tfrac{\eta_2}{d}, \ldots, \tfrac{\eta_2}{d}}_{d}\Big), \\ \sigma \otimes s_{|S|} &= \mathrm{diag}\Big(\underbrace{\tfrac{\zeta_1}{j}, \ldots, \tfrac{\zeta_1}{j}}_{j}, \underbrace{\tfrac{\zeta_2}{j}, \ldots, \tfrac{\zeta_2}{j}}_{j}, \underbrace{0, \ldots, 0}_{2(d-j)}\Big). \end{aligned} \tag{C.5}$$

There are two 'elbows' on $\sigma \otimes s_{|S|}$, at $l = j$ and $l = 2j$. Calculating the ratio of the monotones at these points gives:

$$\frac{V_j\left(\rho \otimes \frac{\mathbb{I}}{d}\right)}{V_j\left(\sigma \otimes s_{|S|}\right)} = \frac{j\frac{\eta_1}{d}}{\zeta_1} = \frac{\eta_1}{\zeta_1} 2^{-S}, \tag{C.6}$$

$$\frac{V_{2j}\left(\rho \otimes \frac{\mathbb{I}}{d}\right)}{V_{2j}\left(\sigma \otimes s_{|S|}\right)} = \begin{cases} 2j\frac{\eta_1}{d} &= 2\eta_1 2^{-S} & \text{if } 2j \leq d, \\ \eta_1 + \frac{2j-d}{d}\eta_2 &= (2\eta_1 - 1) + 2(1-\eta_1)2^{-S} & \text{if } 2j \geq d. \end{cases} \tag{C.7}$$

It is easy to see that $\frac{\eta_1}{\zeta_1} \leq 2\eta_1$ since $\zeta_1 \geq \frac{1}{2}$. Comparing Eq. (C.6) with the second case in Eq. (C.7), it is possible to show that:

$$\frac{V_j\left(\rho \otimes \frac{\mathbb{I}}{d}\right)}{V_j\left(\sigma \otimes s_{|S|}\right)} \leq \frac{V_{2j}\left(\rho \otimes \frac{\mathbb{I}}{d}\right)}{V_{2j}\left(\sigma \otimes s_{|S|}\right)} \Leftrightarrow 2^S \geq \frac{\eta_1 - 2\zeta_1 + 2\eta_1\zeta_1}{2\eta_1\zeta_1 - \zeta_1}. \tag{C.8}$$

As $S \geq 0$, the minimum ratio occurs at $l = j$. Hence:

$$p^*(S) = p^*(0)\, 2^{-S}, \quad S \geq 0. \tag{C.9}$$

Combining these results, we have that for $\eta_1 < \zeta_1$:

$$p^*(S) = \begin{cases} 1 & \text{if } S \leq S_{\rho\to\sigma}, \\ \left(2 - 2^{-S}\right) p^*(0) + \frac{2^{-S}-1}{\zeta_1} & \text{if } S_{\rho\to\sigma} < S \leq 0, \\ p^*(0)\, 2^{-S} & \text{if } 0 < S. \end{cases} \tag{C.10}$$

As an example, in Figure C.1, we plot $p^*(S)$ against $S$ for $\vec{\eta} = \{0.6, 0.4\}$ and $\vec{\zeta} = \{0.85, 0.15\}$.

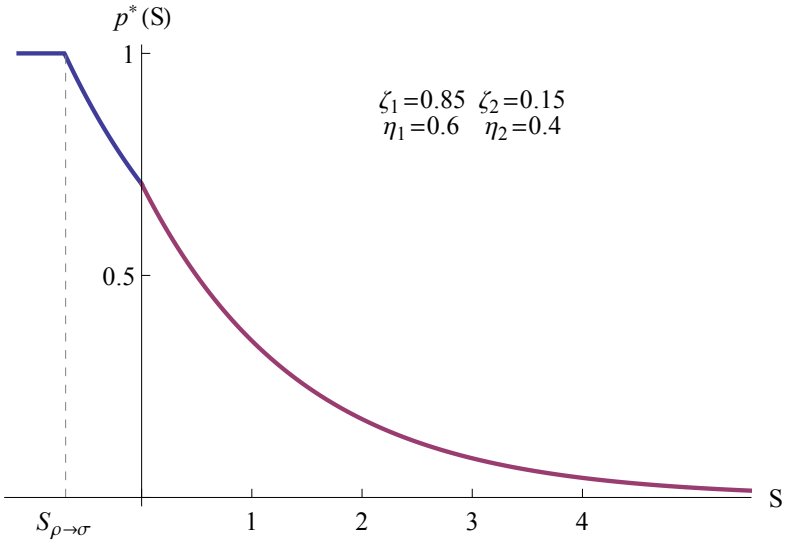



Figure C.1: *Tradeoff between $p^*$ and purity.* Here we show how $p^*$ varies as a function of $S$ for qubits under Noisy Operations when $S_{\rho\to\sigma} < 0$. Note the behavior at $S = 0$, indicating the function is not convex in $S \geq S_{\rho\to\sigma}$.

For completeness, for $\eta_1 \geq \zeta_1$:

$$p^*(S) = \begin{cases} 1 & \text{if } S \leq S_{\rho\to\sigma}, \\ (2\eta_1 - 1) + 2(1-\eta_1)\,2^{-S} & \text{if } S_{\rho\to\sigma} < S \leq \log_2\left(\frac{\eta_1 - 2\zeta_1 + 2\eta_1\zeta_1}{2\eta_1\zeta_1 - \zeta_1}\right), \\ \frac{\eta_1}{\zeta_1}2^{-S} & \text{if } S > \log_2\left(\frac{\eta_1 - 2\zeta_1 + 2\eta_1\zeta_1}{2\eta_1\zeta_1 - \zeta_1}\right). \end{cases} \tag{C.11}$$

# Bibliography


[1] Scott Aaronson, Adam Bouland, Lynn Chua, and George Lowther. $\psi$-epistemic theories: The role of symmetry. *Physical Review A*, 88(3):032111, 2013.

[2] Johan Åberg. Truly work-like work extraction via a single-shot analysis. *Nature communications*, 4, 2013.

[3] Johan Åberg. Catalytic coherence. *Physical Review Letters*, 113(15):150402, 2014.

[4] Álvaro M Alhambra, Jonathan Oppenheim, and Christopher Perry. What is the probability of a thermodynamical transition? *arXiv preprint arXiv:1504.00020*, 2015.

[5] Erika Andersson, Stephen M Barnett, Claire R Gilson, and Kieran Hunter. Minimum-error discrimination between three mirror-symmetric states. *Physical Review A*, 65(5):052308, 2002.

[6] Huzihiro Araki and Elliott H Lieb. Entropy inequalities. *Communications in Mathematical Physics*, 18:160–170, 1970.

[7] Srinivasan Arunachalam, Abel Molina, and Vincent Russo. Quantum hedging in two-round prover-verifier interactions. *arXiv preprint arXiv:1310.7954*, 2013.

[8] Masashi Ban, Keiko Kurokawa, Rei Momose, and Osamu Hirota. Optimum measurements for discrimination among symmetric quantum states and parameter estimation. *International Journal of Theoretical Physics*, 36(6):1269–1288, 1997.

[9] Somshubhro Bandyopadhyay and Rahul Jain. Private communication.

[10] Somshubhro Bandyopadhyay, Rahul Jain, Jonathan Oppenheim, and Christopher Perry. Conclusive exclusion of quantum states. *Physical Review A*, 89(2):022336, 2014.

[11] Ziv Bar-Yossef, TS Jayram, Ravindra Kumar, and D Sivakumar. An information statistics approach to data stream and communication complexity. In *Foundations of Computer Science, 2002. Proceedings. The 43rd Annual IEEE Symposium on*, pages 209–218. IEEE, 2002.

[12] Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. *SIAM Journal on Computing*, 42(3):1327–1363, 2013.

[13] James M Bardeen, Brandon Carter, and Stephen W Hawking. The four laws of black hole mechanics. *Communications in Mathematical Physics*, 31(2):161–170, 1973.

[14] Stephen M Barnett. Minimum-error discrimination between multiply symmetric states. *Physical Review A*, 64(3):030303, 2001.

[15] Stephen M Barnett and Sarah Croke. Quantum state discrimination. *Advances in Optics and Photonics*, 1(2):238–278, 2009.

[16] Howard Barnum and Emanuel Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43(5):2097–2106, 2002.

[17] Jonathan Barrett, Eric G Cavalcanti, Raymond Lal, and Owen JE Maroney. No $\psi$-epistemic model can fully explain the indistinguishability of quantum states. *Physical Review Letters*, 112(25):250403, 2014.

[18] Jonathan Baugh, Osama Moussa, Colm A Ryan, Ashwin Nayak, and Raymond Laflamme. Experimental implementation of heat-bath algorithmic cooling using solid-state nuclear magnetic resonance. *Nature*, 438(7067):470–473, 2005.

[19] Jacob D Bekenstein. Black holes and entropy. *Physical Review D*, 7(8):2333, 1973.

[20] Charles H Bennett. The thermodynamics of computation: A review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

[21] Charles H Bennett. Quantum cryptography using any two nonorthogonal states. *Physical Review Letters*, 68(21):3121, 1992.

[22] Fernando GSL Brandão, Matthias Christandl, and Jon Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 343–352. ACM, 2011.

[23] Fernando GSL Brandão and Gilad Gour. The general structure of quantum resource theories. *arXiv preprint arXiv:1502.03149*, 2015.

[24] Fernando GSL Brandão, Michał Horodecki, Nelly Ng, Jonathan Oppenheim, and Stephanie Wehner. The second laws of quantum thermodynamics. *Proceedings of the National Academy of Sciences*, 112(11):3275–3279, 2015.

[25] Fernando GSL Brandão, Michał Horodecki, Jonathan Oppenheim, Joseph M Renes, and Robert W Spekkens. Resource theory of quantum states out of thermal equilibrium. *Physical Review Letters*, 111(25):250404, 2013.

[26] Mark Braverman. Interactive information complexity. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 505–524. ACM, 2012.

[27] Mark Braverman and Akhila Rao. Information equals amortized communication. *Information Theory, IEEE Transactions on*, 60(10):6058–6069, 2014.

[28] Todd A Brun, Jerry Finkelstein, and N David Mermin. How much state assignments can differ. *Physical Review A*, 65(3):032315, 2002.

[29] Todd A. Brun, Min-Hsiu Hsieh, and Christopher Perry. Compatibility of state assignments and pooling of information. *Phys. Rev. A*, 92:012107, Jul 2015.

[30] Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Reviews of Modern Physics*, 82(1):665, 2010.

[31] Harry Buhrman, Richard Cleve, and Wim Van Dam. Quantum entanglement and communication complexity. *SIAM Journal on Computing*, 30(6):1829–1841, 2001.

[32] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001.

[33] Harry Buhrman, Lukasz Czekaj, Andrzej Grudka, Michał Horodecki, Pawel Horodecki, Marcin Markiewicz, Florian Speelman, and Sergii Strelchuk. Quantum communication complexity advantage implies violation of a Bell inequality. *arXiv preprint arXiv:1502.01058*, 2015.

[34] Harry Buhrman and Ronald de Wolf. Communication complexity lower bounds by polynomials. In *Computational Complexity, 16th Annual IEEE Conference on, 2001.*, pages 120–130. IEEE, 2001.

[35] Harry Buhrman, Wim van Dam, Peter Høyer, and Alain Tapp. Multiparty quantum communication complexity. *Physical Review A*, 60(4):2737, 1999.

[36] Francesco Buscemi and Nilanjana Datta. Distilling entanglement from arbitrary resources. *Journal of Mathematical Physics*, 51(10):102201, 2010.

[37] Francesco Buscemi and Nilanjana Datta. General theory of environment-assisted entanglement distillation. *arXiv preprint arXiv:1009.4464*, 2010.

[38] Francesco Buscemi and Nilanjana Datta. Entanglement cost in practical scenarios. *Physical Review Letters*, 106(13):130503, 2011.

[39] Carlton M Caves, Christopher A Fuchs, and Rüdiger Schack. Conditions for compatibility of quantum-state assignments. *Physical Review A*, 66(6):062111, 2002.

[40] Amit Chakrabart, Yaoyun Shi, Anthony Wirth, and Andrew Yao. Informational complexity and the direct sum problem for simultaneous message complexity. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pages 270–278. IEEE, 2001.

[41] Anthony Chefles. Quantum operations, state transformations and probabilities. *Physical Review A*, 65(5):052314, 2002.

[42] Anthony Chefles and Stephen M Barnett. Strategies for discriminating between non-orthogonal quantum states. *Journal of Modern Optics*, 45(6):1295–1302, 1998.

[43] Chih-Lung Chou. Minimum-error discrimination among mirror-symmetric mixed quantum states. *Physical Review A*, 70(6):062316, 2004.

[44] Chih-Lung Chou and Li-Yi Hsu. Minimum-error discrimination between symmetric mixed quantum states. *Physical Review A*, 68(4):042305, 2003.

[45] Richard Cleve and Harry Buhrman. Substituting quantum entanglement for communication. *Physical Review A*, 56(2):1201, 1997.

[46] Roger Colbeck and Renato Renner. No extension of quantum theory can have improved predictive power. *Nature communications*, 2:411, 2011.

[47] Roger Colbeck and Renato Renner. Is a system's wave function in one-to-one correspondence with its elements of reality? *Physical Review Letters*, 108(15):150402, 2012.

[48] Roger Colbeck and Renato Renner. A system's wave function is uniquely determined by its underlying physical state. *arXiv preprint arXiv:1312.7353*, 2013.

[49] Delphine Collin, Felix Ritort, Christopher Jarzynski, Steven B Smith, Ignacio Tinoco, and Carlos Bustamante. Verification of the crooks fluctuation theorem and recovery of RNA folding free energies. *Nature*, 437(7056):231–234, 2005.

[50] Robert J Collins, Ross J Donaldson, Vedran Dunjko, Petros Wallden, Patrick J Clarke, Erika Andersson, John Jeffers, and Gerald S Buller. Realization of quantum digital signatures without the requirement of quantum memory. *Physical Review Letters*, 113(4):040502, 2014.

[51] Gavin E Crooks. Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems. *Journal of Statistical Physics*, 90(5-6):1481–1487, 1998.

[52] Piotr Ćwikliński, Michał Studziński, Michał Horodecki, and Jonathan Oppenheim. Towards fully quantum second laws of thermodynamics: limitations on the evolution of quantum coherences. *arXiv preprint arXiv:1405.5029*, 2014.

[53] Oscar CO Dahlsten, Mahn-Soo Choi, Daniel Braun, Andrew JP Garner, Nicole Yunger Halpern, and Vlatko Vedral. Equality for worst-case work at any protocol speed. *arXiv preprint arXiv:1504.05152*, 2015.

[54] Oscar CO Dahlsten, Renato Renner, Elisabeth Rieper, and Vlatko Vedral. Inadequacy of von Neumann entropy for characterizing extractable work. *New Journal of Physics*, 13(5):053015, 2011.

[55] Ronald de Wolf. Nondeterministic quantum query and communication complexities. *SIAM Journal on Computing*, 32(3):681–699, 2003.

[56] Lidia del Rio, Johan Åberg, Renato Renner, Oscar Dahlsten, and Vlatko Vedral. The thermodynamic meaning of negative entropy. *Nature*, 474(7349):61–63, 2011.

[57] Dennis Dieks. Overlap and distinguishability of quantum states. *Physics Letters A*, 126(5):303–306, 1988.

[58] Andrew C Doherty, Pablo A Parrilo, and Federico M Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2):022308, 2004.

[59] Runyao Duan, Simone Severini, and Andreas Winter. Zero-error communication via quantum channels, noncommutative graphs, and a quantum Lovász number. *Information Theory, IEEE Transactions on*, 59(2):1164–1174, 2013.

[60] Runyao Duan, Simone Severini, and Andreas Winter. On zero-error communication via quantum channels in the presence of noiseless feedback. *arXiv preprint arXiv:1502.02987*, 2015.

[61] Dario Egloff, Oscar CO Dahlsten, Renato Renner, and Vlatko Vedral. Laws of thermodynamics beyond the von Neumann regime. *arXiv preprint arXiv:1207.0434*, 2012.

[62] Yonina C Eldar and G David Forney Jr. On quantum detection and the square-root measurement. *Information Theory, IEEE Transactions on*, 47(3):858–872, 2001.

[63] Yonina C Eldar, Alexandre Megretski, and George C Verghese. Designing optimal quantum detectors via semidefinite programming. *Information Theory, IEEE Transactions on*, 49(4):1007–1012, 2003.

[64] Yonina C Eldar, Alexandre Megretski, and George C Verghese. Optimal detection of symmetric mixed quantum states. *Information Theory, IEEE Transactions on*, 50(6):1198–1207, 2004.

[65] Philippe Faist, Frédéric Dupuis, Jonathan Oppenheim, and Renato Renner. The minimal work cost of information processing. *Nature communications*, 6, 2015.

[66] LP Faucheux, LS Bourdieu, PD Kaplan, and AJ Libchaber. Optical thermal ratchet. *Physical Review Letters*, 74(9):1504, 1995.

[67] Jaromír Fiurášek and Miroslav Ježek. Optimal discrimination of mixed quantum states involving inconclusive results. *Physical Review A*, 67(1):012321, 2003.

[68] Jörg Flum and Martin Grohe. Parameterized complexity theory, volume XIV of texts in theoretical computer science. An EATCS series, 2006.

[69] Rodrigo Gallego and Leandro Aolita. The resource theory of steering. *arXiv preprint arXiv:1409.5804*, 2014.

[70] Rodrigo Gallego, Jens Eisert, and Henrik Wilming. Defining work from operational principles. *arXiv preprint arXiv:1504.05056*, 2015.

[71] Ernesto F Galvao and Lucien Hardy. Substituting a qubit for an arbitrarily large number of classical bits. *Physical Review Letters*, 90(8):087902, 2003.

[72] Anat Ganor, Gillat Kol, and Ran Raz. Exponential separation of information and communication. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 176–185. IEEE, 2014.

[73] Dmitry Gavinsky. Classical interaction cannot replace quantum nonlocality. *arXiv preprint arXiv:0901.0956*, 2009.

[74] Dmitry Gavinsky, Julia Kempe, and Ronald de Wolf. Strengths and weaknesses of quantum fingerprinting. In *Computational Complexity, 2006. CCC 2006. Twenty-First Annual IEEE Conference on*, pages 8–pp. IEEE, 2006.

[75] Dmitry Gavinsky, Julia Kempe, Oded Regev, and Ronald de Wolf. Bounded-error quantum state identification and exponential separations in communication complexity. *SIAM Journal on Computing*, 39(1):1–24, 2009.

[76] John Goold, Marcus Huber, Arnau Riera, Lídia del Rio, and Paul Skrzypzyk. The role of quantum information in thermodynamics—a topical review. *arXiv preprint arXiv:1505.07835*, 2015.

[77] Gilad Gour, Markus P Müller, Varun Narasimhachar, Robert W Spekkens, and Nicole Yunger Halpern. The resource theory of informational nonequilibrium in thermodynamics. *Physics Reports*, 2015.

[78] Gilad Gour and Robert W Spekkens. The resource theory of quantum reference frames: manipulations and monotones. *New Journal of Physics*, 10(3):033023, 2008.

[79] Nicole Yunger Halpern, Andrew JP Garner, Oscar CO Dahlsten, and Vlatko Vedral. Unification of fluctuation theorems and one-shot statistical mechanics. *arXiv preprint arXiv:1409.3878*, 2014.

[80] Godfrey Harold Hardy, John Edensor Littlewood, and George Pólya. *Inequalities*. Cambridge University Press, 1952.

[81] Lucien Hardy. Are quantum states real? *International Journal of Modern Physics B*, 27(01n03):1345012, 2013.

[82] Nicholas Harrigan and Robert W Spekkens. Einstein, incompleteness, and the epistemic view of quantum states. *Foundations of Physics*, 40(2):125–157, 2010.

[83] Masahito Hayashi, Akinori Kawachi, and Hirotada Kobayashi. Quantum measurements for hidden subgroup problems with optimal sample complexity. *arXiv preprint quant-ph/0604174*, 2006.

[84] Carl W Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1(2):231–252, 1969.

[85] Alexander S Holevo. Statistical decision theory for quantum systems. *Journal of Multivariate Analysis*, 3(4):337–394, 1973.

[86] Alfred Horn. Doubly stochastic matrices and the diagonal of a rotation matrix. *American Journal of Mathematics*, pages 620–630, 1954.

[87] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge University Press, 2012.

[88] Michał Horodecki, Paweł Horodecki, and Jonathan Oppenheim. Reversible transformations from pure to mixed states and the unique measure of information. *Physical Review A*, 67(6):062104, 2003.

[89] Michał Horodecki and Jonathan Oppenheim. Fundamental limitations for quantum and nanoscale thermodynamics. *Nature communications*, 4, 2013.

[90] Michał Horodecki and Jonathan Oppenheim. (Quantumness in the context of) Resource theories. *International Journal of Modern Physics B*, 27(01n03):1345019, 2013.

[91] Igor D Ivanovic. How to differentiate between non-orthogonal states. *Physics Letters A*, 123(6):257–259, 1987.

[92] Rahul Jain, Hartmut Klauck, and Ashwin Nayak. Direct product theorems for classical communication complexity via subdistribution bounds. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 599–608. ACM, 2008.

[93] Rahul Jain and Amiya Nayak. The space complexity of recognizing well-parenthesized expressions in the streaming model: The index function revisited. *Information Theory, IEEE Transactions on*, 60(10):6646–6668, 2014.

[94] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A direct sum theorem in communication complexity via message compression. *Lecture notes in computer science*, 2719:300–315, 2003.

[95] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of set disjointness. In *Proceedings-Annual Symposium on Foundations of Computer Science*, pages 220–229. IEEE, 2003.

[96] Dominik Janzing, Pawel Wocjan, Robert Zeier, Rubino Geiss, and Th Beth. Thermodynamic cost of reliability and low temperatures: Tightening Landauer's principle and the second law. *International Journal of Theoretical Physics*, 39(12):2717–2753, 2000.

[97] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[98] M Ježek, J Řeháček, and J Fiurášek. Finding optimal strategies for minimum-error quantum-state discrimination. *Physical Review A*, 65(6):060301, 2002.

[99] Daniel Jonathan and Martin B Plenio. Entanglement-assisted local manipulation of pure quantum states. *Physical Review Letters*, 83(17):3566, 1999.

[100] Daniel Jonathan and Martin B Plenio. Minimal conditions for local pure-state entanglement manipulation. *Physical Review Letters*, 83(7):1455, 1999.

[101] Richard Jozsa and Jürgen Schlienz. Distinguishability of states and von Neumann entropy. *Physical Review A*, 62(1):012301, 2000.

[102] Iordanis Kerenidis, Sophie Laplante, Virginie Lerays, Jérémie Roland, and David Xiao. Lower bounds on information complexity via zero-communication protocols and applications. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on*, pages 500–509. IEEE, 2012.

[103] Alexei Kitaev. Private communication with Todd A. Brun and Min-Hsiu Hsieh.

[104] Donald E Knuth. Big omicron and big omega and big theta. *ACM Sigact News*, 8(2):18–24, 1976.

[105] Kamil Korzekwa, Matteo Lostaglio, Jonathan Oppenheim, and David Jennings. The extraction of work from quantum coherence. *arXiv preprint arXiv:1506.07875*, 2015.

[106] Robert L Kosut, Ian Walmsley, Yonina Eldar, and Herschel Rabitz. Quantum state detector design: Optimal worst-case a posteriori performance. *arXiv preprint quant-ph/0403150*, 2004.

[107] Ilan Kremer. Quantum communication. Master's thesis, The Hebrew University of Jerusalem, 1995.

[108] Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

[109] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM journal of research and development*, 5(3):183–191, 1961.

[110] MS Leifer. Is the quantum state real? A review of $\psi$-ontology theorems. *arXiv preprint arXiv:1409.1570*, 2014.

[111] Peter G Lewis, David Jennings, Jonathan Barrett, and Terry Rudolph. Distinct quantum states can be compatible with a single state of reality. *Physical Review Letters*, 109(15):150404, 2012.

[112] Noah Linden, Sandu Popescu, and Paul Skrzypczyk. How small can thermal machines be? The smallest possible refrigerator. *Physical Review Letters*, 105(13):130401, 2010.

[113] Zi-Wen Liu, Christopher Perry, Yechao Zhu, Dax Enshan Koh, and Scott Aaronson. Doubly infinite separation of quantum information and communication. *arXiv preprint arXiv:1507.03546*, 2015.

[114] Johan Löfberg. YALMIP: A toolbox for modeling and optimization in MATLAB. In *Computer Aided Control Systems Design, 2004 IEEE International Symposium on*, pages 284–289. IEEE, 2004.

[115] Matteo Lostaglio, David Jennings, and Terry Rudolph. Description of quantum coherence in thermodynamic processes requires constraints beyond free energy. *Nature communications*, 6, 2015.

[116] Matteo Lostaglio, Kamil Korzekwa, David Jennings, and Terry Rudolph. Quantum coherence, time-translation symmetry, and thermodynamics. *Physical Review X*, 5(2):021001, 2015.

[117] Matteo Lostaglio, Markus P Mueller, and Michele Pastena. Extracting work from absence of correlations. *arXiv preprint arXiv:1409.3258*, 2014.

[118] Maria Manosas, Alessandro Mossa, Nuria Forns, Josep Maria Huguet, and Felix Ritort. Dynamic force spectroscopy of DNA hairpins: II. Irreversibility and dissipation. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(02):P02061, 2009.

[119] Owen JE Maroney. How statistical are quantum states? *arXiv preprint arXiv:1207.6906*, 2012.

[120] Albert W Marshall, Ingram Olkin, and Barry Arnold. *Inequalities: theory of majorization and its applications*. Springer Science & Business Media, 2010.

[121] Lluis Masanes and Jonathan Oppenheim. A derivation (and quantification) of the third law of thermodynamics. *arXiv preprint arXiv:1412.3828*, 2014.

[122] Serge Massar, Dave Bacon, Nicolas J Cerf, and Richard Cleve. Classical simulation of quantum entanglement without local hidden variables. *Physical Review A*, 63(5):052305, 2001.

[123] Abel Molina and John Watrous. Hedging bets with correlated quantum strategies. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, page rspa20110621. The Royal Society, 2012.

[124] Ashley Montanaro. On the distinguishability of random quantum states. *Communications in Mathematical Physics*, 273(3):619–636, 2007.

[125] Ashley Montanaro. A lower bound on the probability of error in quantum state discrimination. In *Information Theory Workshop, 2008. ITW'08. IEEE*, pages 378–380. IEEE, 2008.

[126] Ashley Montanaro and Andreas Winter. A lower bound on entanglement-assisted quantum communication complexity. In *Automata, Languages and Programming*, pages 122–133. Springer, 2007.

[127] Alessandro Mossa, Maria Manosas, Nuria Forns, Josep Maria Huguet, and Felix Ritort. Dynamic force spectroscopy of DNA hairpins: I. Force kinetics and free energy landscapes. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(02):P02060, 2009.

[128] Robert Franklin Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of $n$ letters. *Proceedings of the Edinburgh Mathematical Society*, 21:144–162, 1902.

[129] Kenji Nakahira and Tsuyoshi Sasaki Usuda. Minimum-Bayes-cost discrimination for symmetric quantum states. *Physical Review A*, 86(6):062305, 2012.

[130] Miguel Navascués and Luis Pedro García-Pintos. Non-thermal quantum channels as a thermodynamical resource. *arXiv preprint arXiv:1501.02597*, 2015.

[131] Miguel Navascués and Sandu Popescu. How energy conservation limits our measurements. *Physical Review Letters*, 112(14):140502, 2014.

[132] Nelly Ng, Laura Mančinska, Cristina Cirstoiu, Jens Eisert, and Stephanie Wehner. Limits to catalysis in quantum thermodynamics. *arXiv preprint arXiv:1405.3039*, 2014.

[133] Michael A Nielsen. Conditions for a class of entanglement transformations. *Physical Review Letters*, 83(2):436, 1999.

[134] Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge University Press, 2010.

[135] Tomohiro Ogawa and Hiroshi Nagaoka. Strong converse to the quantum channel coding theorem. *arXiv preprint quant-ph/9808063*, 1998.

[136] Rudolf Ernst Peierls. *More surprises in theoretical physics*, volume 19. Princeton University Press, 1991.

[137] Asher Peres. How to differentiate between non-orthogonal states. *Physics Letters A*, 128(1):19, 1988.

[138] Asher Peres. *Quantum theory: Concepts and methods*, volume 57. Springer Science & Business Media, 1995.

[139] Christopher Perry, Piotr Ćwikliński, Janet Anders, Michał Horodecki, and Jonathan Oppenheim. A sufficient set of experimentally implementable thermal operations. *arXiv preprint arXiv:1511.06553*, 2015.

[140] Christopher Perry, Rahul Jain, and Jonathan Oppenheim. Communication tasks with infinite quantum-classical separation. *Physical Review Letters*, 115:030504, Jul 2015.

[141] Sandu Popescu. Maximally efficient quantum thermal machines: The basic principles. *arXiv preprint arXiv:1009.2536*, 2010.

[142] Matthew F Pusey, Jonathan Barrett, and Terry Rudolph. On the reality of the quantum state. *Nature Physics*, 8(6):475–478, 2012.

[143] Daowen Qiu. Minimum-error discrimination between mixed quantum states. *Physical Review A*, 77(1):012328, 2008.

[144] Daowen Qiu and Lvjun Li. Minimum-error discrimination of quantum states: Bounds and comparisons. *Physical Review A*, 81(4):042329, 2010.

[145] Ran Raz. Exponential separation of quantum and classical communication complexity. In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, pages 358–367. ACM, 1999.

[146] David Reeb and Michael M Wolf. An improved Landauer principle with finite-size corrections. *New Journal of Physics*, 16(10):103011, 2014.

[147] Oded Regev and Bo'az Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 31–40. ACM, 2011.

[148] Joseph M Renes. Work cost of thermal operations in quantum thermodynamics. *The European Physical Journal Plus*, 129(7):1–7, 2014.

[149] Joseph M Renes. Relative submajorization and its use in quantum resource theories. *arXiv preprint arXiv:1510.03695*, 2015.

[150] Terry Rudolph and Robert W Spekkens. Quantum state targeting. *Physical Review A*, 70(5):052306, 2004.

[151] Sina Salek and Karoline Wiesner. Fluctuations in single-shot $\epsilon$-deterministic work extraction. *arXiv preprint arXiv:1504.05111*, 2015.

[152] Issai Schur. Uber eine klasse von mittelbildungen mit anwendungen auf die determinantentheorie. *Sitzungsberichte der Berliner Mathematischen Gesellschaft*, 22:9–20, 1923.

[153] HED Scovil and EO Schulz-DuBois. Three-level masers as heat engines. *Physical Review Letters*, 2(6):262, 1959.

[154] Marlan O Scully. Extracting work from a single thermal bath via quantum negentropy. *Physical Review Letters*, 87(22):220601, 2001.

[155] Viviana Serreli, Chin-Fa Lee, Euan R Kay, and David A Leigh. A molecular information ratchet. *Nature*, 445(7127):523–527, 2007.

[156] Paul Skrzypczyk, Nicolas Brunner, Noah Linden, and Sandu Popescu. The smallest refrigerators can reach maximal efficiency. *arXiv preprint arXiv:1009.0865*, 2010.

[157] Paul Skrzypczyk, Anthony J Short, and Sandu Popescu. Work extraction and thermodynamics for individual quantum systems. *Nature communications*, 5, 2014.

[158] Robert W Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A*, 71(5):052108, 2005.

[159] Robert W Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75(3):032110, 2007.

[160] Jos F Sturm. Using SeDuMi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software*, 11(1-4):625–653, 1999.

[161] Leo Szilard. Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen. *Zeitschrift für Physik*, 53(11-12):840–856, 1929.

[162] Dave Touchette. Quantum information complexity and amortized communication. *arXiv preprint arXiv:1404.3733*, 2014.

[163] Wim van Dam and Patrick Hayden. Universal entanglement transformations without communication. *Physical Review A*, 67(6):060302, 2003.

[164] Guifré Vidal. Entanglement of pure states for a single copy. *Physical Review Letters*, 83(5):1046, 1999.

[165] Guifré Vidal. Entanglement monotones. *Journal of Modern Optics*, 47(2-3):355–376, 2000.

[166] John Watrous. *Lecture notes, CS 766/QIC 820 Theory of Quantum Information, University of Waterloo. See lecture 7, Semidefinite programming, and Lecture 8, Semidefinite Programs for Fidelity and Optimal Measurements.* Fall, 2011.

[167] Stephanie Wehner. Tsirelson bounds for generalized Clauser-Horne-Shimony-Holt inequalities. *Physical Review A*, 73(2):022110, 2006.

[168] Hermann Weyl. Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Mathematische Annalen*, 71(4):441–479, 1912.

[169] Mark M Wilde. *Quantum information theory.* Cambridge University Press, 2013.

[170] Andreas Winter and Dong Yang. Operational resource theory of coherence. *arXiv preprint arXiv:1506.07975*, 2015.

[171] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: Theory, algorithms, and applications*, volume 27. Springer Science & Business Media, 2000.

[172] Mischa P Woods, Nelly Ng, and Stephanie Wehner. The maximum efficiency of nano heat engines depends on more than temperature. *arXiv preprint arXiv:1506.02322*, 2015.

[173] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213. ACM, 1979.

[174] Horace P Yuen, Robert S Kennedy, and Melvin Lax. Optimum testing of multiple hypotheses in quantum detection theory. *Information Theory, IEEE Transactions on*, 21(2):125–134, 1975.

[175] Chuan-Wei Zhang, Chuan-Feng Li, and Guang-Can Guo. General strategies for discrimination of quantum states. *Physics Letters A*, 261(1):25–29, 1999.